DOCUMENT RESUME

ED 022 703 24 SE 005 193

By-Brownell, William A.; And Others
ARITHMETICAL COMPUTATION: COMPETENCE AFTER THREE YEARS OF LEARNING UNDER DIFFERING INSTRUCTIONAL PROGRAMS.
California Univ., Berkeley.
Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Research.
Bureau No-BR-6-1077
Note-255p.
EDRS Price MF-$1.00 HC-$10.28
Descriptors-ACHIEVEMENT, *ARITHMETIC, *COMPARATIVE ANALYSIS, *INSTRUCTION, LEARNING, MATHEMATICS, TEACHING TECHNIQUES
Identifiers-Scotland

Reported are the results and conclusions of an arithmetic investigation made in the schools of Scotland in the spring and fall of 1966. The first problem in this investigation was to ascertain which, if either, of two unlike programs of instruction was more effective in developing skill in computation. The second was to determine the value of an unusual design for this kind of evaluative research. The programs in question are the Cuisenaire (Cui.) system of instruction and the Traditional (Tra.) system, both systems defined as they were taught in the school year 1963-65. The subjects for this study were 1109 Scottish children ranging in age from 91 to 119 months, with a mean age of approximately 100 months. One group of subjects consisted of 539 children who had studied arithmetic according to the Cui. program for three years, in Primary I, II, and III. The second group of 570 children had been taught arithmetic in the same grades according to the Tra. program. From the evidence obtained, it is reasonable to suggest that children identified as low in intelligence and exposed to a relatively long period of instruction in arithmetic will gain more through involvement in the Cui. program. No claims for the superiority of either program can be made with respect to other sub-samples or the total samples selected for this investigation. (RP)

BR-6-1077
PA-24

ARITHMETICAL COMPUTATION: COMPETENCE AFTER THREE YEARS OF LEARNING UNDER DIFFERING INSTRUCTIONAL PROGRAMS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

by

William A. Brownell
University of California, Berkeley

with the cooperation of

Arden K. Ruddell
University of California

and with the assistance of

Margaret L. F. Law and Robert J. Allan
Moray House College of Education
Edinburgh, Scotland

ED022703

SE005 193

ARITHMETICAL COMPUTATION: COMPETENCE AFTER THREE YEARS OF LEARNING UNDER DIFFERING INSTRUCTIONAL PROGRAMS

by

William A. Brownell
University of California, Berkeley

with the cooperation of

Arden K. Ruddell
University of California

and with the assistance of

Margaret L. F. Law and Robert J. Allan
Moray House College of Education
Edinburgh, Scotland

ACKNOWLEDGMENTS

I wish to thank the U. S. Office of Education of the U. S. Department of Health, Education, and Welfare for funds to support the research reported in this manuscript, and the University of California, Berkeley, for assisting me in procuring these funds and then managing them.

Next, I should express my gratitude to those in Scotland who so generously--so magnificently--cooperated in bringing the investigation to a successful conclusion, more particularly

- - - - to Miss Margaret L. F. Law, Senior Lecturer in Methods and Mr. Robert J. Allan, Senior Lecturer in Mathematics, both of Moray House College of Education, Edinburgh, for serving as chairmen of the two panels that constructed the arithmetic tests, for finding schools for the testing, for writing statements concerning the two arithmetic programs under study, for identifying persons to administer and score the tests, and for wise counsel throughout the course of the inquiry; and

- - - - to Mr. James Carr, Head Master of the West School, Miss Margaret Leitch, Infant Mistress in the same school, Mr. John Mitchell, Head Master of the Groathill School, and Mrs. David Younger, teacher in the Valley School, who, as members of the two panels, made contributions of significant value to the inquiry; and

- - - - to the Directors of Education, their Head Masters and Head Mistresses, and their classroom teachers, who furnished subjects for the study; and, perhaps most of all, to the hundreds of children who patiently submitted to hours of testing;

- - - - to Dr. D. A. Walker, Director of the Scottish Council for Educational Research, Edinburgh, his assistant, Miss Jessie Gray, and other members of the Council staff for help of many kinds, among them the making of arrangements for working space for me in their office building; and

- - - - to the eleven charming, intelligent, and responsible young ladies, all members of the graduating class at Moray

House College of Education, who cheerfully did all the testing and the scoring of tests under a most demanding schedule.

Last of all, I am very deeply indebted to my colleague, Dr. Arden K. Ruddell, of the School of Education, at the University of California, Berkeley, who directed the statistical work reported in Chapter III and wrote that chapter and part of the final chapter, who read the entire manuscript and helped to improve its completeness and clarity, and who always found time to consult with me and to help me out of my troubles.

William A. Brownell

CONTENTS

Page

Acknowledgements ii

Part I. Purpose and Procedure

Chapter I. The Problem and General Plan of Attack 1

Skill in computation as an objective of arithmetic instruction 1

Limited scope of the investigation.. 4

The general plan of attack.......... 6

The program compared................ 9

Chapter II. Details of the Procedure............ 26

The subjects........................ 26

Selection of schools................ 28

The tests........................... 29

The testing......................... 36

Organization of the following chapters........................ 38

Part II. An Examination of the Results by Analysis of Variance

Chapter III. Main Effects and Interaction of Variables in this Investigation..... 41

Introduction........................ 41

Total group results--a first impression......................... 42

Sub-sample performance--a second impression......................... 46

Concluding Statement................ 62

CONTENTS (Continued)

Page

Part III. Relative Achievement and Relative Extent of Transfer, According to Type of Skill

Chapter IV. Purpose and Procedure 69

Program samples......................... 69

Method of analysis...................... 71

Chapter V. Relative Achievement on Skills Taught in Both Programs................... 79

Set 1. The simple number combinations................... 79

Set 2. Supplying numerals in number sentences...................... 87

Set 3. Horizontal and/or vertical addition....................... 93

Set 4. Fractions..................... 96

Set 5. Structure and divisibility of numbers..................... 99

Set 6. Two-operation examples......... 103

Set 7. Time measures; Scottish money.. 105

Summary and interpretation............. 107

Concluding statement................... 125

Chapter VI. Extent of Transfer: The Tra. Subjects... 127

Set 1. One-operation items........... 130

Set 2. Fraction computations; two-operation examples with whole numbers and fractions.......... 133

Set 3. Computation with one and two brackets, fractions and whole numbers................. 138

CONTENTS (Continued)

Page

Set 4. Number meanings; doubling numbers............... 141

Set 5. Number progressions; relative size of numbers....... 143

Set 6. Two- and three-operation examples with one or two sets of brackets............... 145

Set 7. Computation with Scottish money, all with fractions...... 148

Set 8. Measurement examples, plus two miscellaneous items........ 150

Concluding statement.................. 152

Chapter VII. Extent of Transfer: The Cui. Subjects 157

Set 1. Column addition............... 159

Set 2. Horizontal addition........... 165

Set 3. Subtraction................... 167

Set 4. Division...................... 173

Set 5. Multiplication................ 178

Set 6. The meaning of numbers......... 181

Set 7. Completing two-operation number sentences............... 184

Concluding statement.................. 188

Part IV. Conclusions and Implications

Chapter VIII. Summary and Findings, with Related Research.................. 196

Purpose of the study.................. 196

CONTENTS (Continued)

Page

Procedure.............................. 197

Summary of the findings: relative achievement.................... 202

Summary of the findings: transfer of learning.................... 209

Review of related research............. 222

Critique of the design of the investigation...................... 230

Appendix.. 235

PART I

PURPOSE AND PROCEDURE

CHAPTER I

THE PROBLEM AND GENERAL PLAN OF ATTACK

In this report are presented the results and conclusions of an arithmetic investigation made in the schools of Scotland in the spring and fall of 1966. The purpose of the inquiry was two-fold. The first problem was to ascertain which, if either, of two unlike programs of instruction was the more effective in developing skill in computation. The second was to determine the value of an unusual design for this kind of evaluative research.

Skill in Computation as an Objective of Arithmetic Instruction

In American schools, as in schools generally throughout the world, computational skill has always been regarded as one of the aims of arithmetic instruction. True, in this country the importance attached to such skill has seemed to vary considerably in this century. At times it has appeared to be the principal aim, as in the 20's and 30's. Most children in this period were subjected daily to rigorous drill in little else than calculating answers for endless sets of abstract examples.

On the other hand at times, also in the 30's and in the early years of the 40's, in many classrooms proficiency in computation seemed to be minimized as an objective of arithmetic instruction. In such schools it was supposed that children would inevitably acquire computational skill as they dealt with the quantitative situations they encountered more or less incidentally in the course of their daily lives. Hence, it was thought, there was no need for regularly organized and systematic practice specifically designed to produce computational proficiency. Under this regimen, it will be noted, skill in computation nevertheless remained one of the purposes of instruction and was not totally neglected.

The difference between the two extreme programs of teaching cited turn out to be less one of ends than one of means, the goal in both being substantially the same.

There is of course good reason why computational skill has been consistently viewed as an essential learning outcome in arithmetic. Without this skill, and in the absence of computers and like apparatus, one is helpless in any civilized community, for one is constantly beset by occasions calling for the ability to interpret quantitative relationships economically, accurately, and efficiently.

Nevertheless, the ability just referred to is not equivalent to computational expertness alone. A child can compute well with abstract numbers and be at a loss in solving problems, both those of the arithmetic textbook and those arising in practical living outside of school. He may know how quickly to obtain the correct answer for 3 x 18, but may not know how to solve the problem, "How many peaches are there in three bags if there are eighteen peaches in each?" For he cannot identify in the verbal statement the need for multiplication: the cue "x" is absent. Told by his teacher that he must multiply, he readily obtains the required 54: the word "multiply," like "x" is to him a familiar cue to multiplication. But in no sense of the word has he solved a problem--except that of getting himself out of a despairful situation.

Neither child nor adult computes for the sake of computing. He computes to achieve a purpose outside itself as a means of extricating himself from a quantitative predicament. Computational proficiency completely divorced from ability to use that proficiency is scarcely worth developing. Yet, this proficiency cannot be developed through applications, and for at least two reasons. First, the meanings requisite to functional computation reside, not in the social aspects of quantitative situations, but in mathematics. Second, knowledge of the rationale of computation must be supplemented by practice in computation, separate from applications, in order to assure lasting facility in calculations with numbers.[1]

---

[1] To complete the picture: since skill in computation does not automatically and fully carry over to skill in solving problems (of whatever kind), instruction in computation has to be complemented by instruction of quite a different

Limited Scope of the Investigation

The final test of the worth of a system of arithmetic instruction is its effectiveness in enabling children as children, and later as adults, to live efficiently, intelligently, and richly in their quantitative culture. The multiplicity of skills, concepts, attitudes, and the like which comprise the sort of effectiveness described means that the evaluation of a single program, to say nothing of two or more programs, is an exceedingly complicated enterprise, so complicated indeed that comprehensive evaluations are unlikely to be made.

Instead, customarily the scope of evaluative studies is restricted. Data are obtained in the classroom, not in out-of-school situations, and they are obtained from children aged 6 to 14, and not from adults. The critical measures procured are scores on paper-and-pencil tests of computational skill and/or of skill in solving artificial verbal problems like those in textbooks. Inquiries of this character can be

---

kind. Problems, even when they arise in connection with concrete objects, are soon or late presented and solved with verbal symbols. Hence, difficulty of language interpretation occurs, and there have to be numerous and varied experiences in translating the language of problems into recognized needs for this or that mathematical operation. Inability to perform this last task means that proficiency in abstract computation has no chance to function.

of value, but are far from being the kind we should ultimately seek to make.

Let it be said now at the outset that the present investigation is limited in scope. Paper-and-pencil tests, and these only were used in Scottish schools. (More crucial evidence than test scores might have been secured by observing and questioning children at work; but these research techniques were not practicable in this instance.) The tests given had to do with computation, and this alone. For reasons which will appear later, the subjects had to be children just starting Primary IV (not identical with Grade 4 in American schools), whereas tests administered at a later point in schooling would seem to have been preferable.

Not all the limitations have been mentioned, but enough have been cited to indicate that caution is mandatory in interpreting the findings. It should be obvious that neither of the two programs studied can be said to have been evaluated as a whole and that comparisons of their relative effectiveness must be restricted to computational proficiency. There is little basis, therefore, for concluding, or for even assuming, that the program found to be superior in engendering computational skill must therefore be preferred in all respects to its rival.

The limitations just discussed relate only to one of the purposes of this investigation; namely, that of comparing two programs of instruction for their effectiveness in promoting computational skill. They do not relate to the second purpose, which was to submit to trial a new (as far as is known) design for evaluating systems of instruction in a given subject matter area.

The General Plan of Attack

In the typical study set up to evaluate by tests differing programs of teaching a school subject, errors are often committed in selecting the instruments used. To illustrate: about five years ago an investigator undertook to compare the results of teaching about the same two arithmetic programs examined in the present inquiry. One of these programs will be designated here as the Traditional (or Tra.) Program. The term "traditional" is employed in no derogatory sense, to mean "old fashioned" or "out-moded." Rather, it refers to a system which, dating back at least fifty years, had evolved by 1960 through the adoption of relatively minor changes, with the emphasis still placed primarily on formal computation and problem solving and with comparatively little of the new content and organization characteristic of "modern mathematics." The second program in this study was the Cuisenaire (or Cui.) Program, well known if not extensively used in American schools. (See fuller accounts of the two programs in the section below.)

To secure measures of attainment the investigator in question chose a highly respected standard test battery.

Unfortunately, the battery, as far as arithmetic is concerned, is best adapted to the survey function of measurement and is not suitable to the uses to which it was put. Moreover, the arithmetic tests in the battery were based upon the "traditional" approach to teaching this subject. Therefore, the children in Tra. schools had what was certainly an unintended advantage in that they were tested upon what they had been taught. On the other hand, the Cui. children were tested in part at least on material they had not been taught and were not tested on many skills they had learned as features of their particular program. It is small wonder that, in general, the Tra. subjects outscored the Cui.subjects.

From his data the investigator concluded that the Tra. system of instruction is to be preferred to the Cui. This judgment can be valid only on the assumption that the content of the tests used represents precisely what should be taught in arithmetic, an assumption which was, and is, certainly debatable.

In the present research an attempt was made to avoid prejudicing the case either for the Tra. or for the Cui. program. The aim was to employ a testing program that was comprehensive and fair to both groups of subjects. The arithmetic tests used, of which there were three, were devised by Scottish educators for Scottish school children. (Details of the procedure are postponed to Chapter II.)

Suffice it to say here that

1. A Common Test in computation was prepared which contained only computational items acceptable to the members of two panels, one consisting of experts in the Cui. program and the other, of experts in the Tra. program.

2. A special Cui. Test constructed by the Cui. panel, was made up of items peculiar to the Cui. program and hence not useable in the Common Test, and

3. A corresponding special Tra. Test was devised by the Tra. panel for their program and contained only items not taught in the Cui. program and so, not included in the Common Test.

All children, whether taught according to the Cui. program or according to the Tra. program, took all three arithmetic tests (along with a test of "intelligence" or "scholastic aptitude"). As a consequence

1. the Common Test was actually an achievement test for both groups, for both were tested on items known to have been taught them. Theoretically, comparisons of scores of the two groups should provide a means of determining the relative effectiveness of the two programs in engendering the same computational skills;
2. the Cui. Test, an achievement test for the Cui. pupils, was quite largely a transfer test for the Tra. pupils and (theoretically again) should show the extent to which the Tra. program had developed understandings

and insights that could be carried over to untaught computational skills, and

3. by the same token, the Tra. Test, an achievement test for the Tra. pupils, served as a transfer test for the Cui. pupils.

The Programs Compared

It did not seem advisable for an outsider, like the investigator, to formulate descriptions of the Cui. and of the Tra. programs as taught in Scottish schools. Therefore, the chairmen of the two panels of experts were invited, and graciously consented, to take this responsibility. The statement concerning the Tra. program was written by Mr. Robert J. Allan, Senior Lecturer in Mathematics in the Moray House College of Education in Edinburgh; that for the Cui. Program, by Miss Margaret L. F. Law, Senior Lecturer in Methods in the same institution. Both contributors were asked to say something about recent and current developments in arithmetic education, to indicate the place (the relative importance) of computation in the programs, and to describe, with some details, what is done in teaching computational skills in the first three years of schooling. Their statements follow.

Computation in the Cuisenaire Program

(Miss Margaret L. F. Law)

During the last five years there have probably been more radical changes in educational policies and practices in Scotland than took place during the twenty years preceding the nineteen-sixties. All aspects of child learning have been under review, resulting in certain areas in a pronounced swing away from formal instruction and sometimes in perhaps rather indiscriminate acceptance of 'discovery' and 'activity' methods. Extremists would have schools abandon time-tables, programmes, and all directed lessons, believing rather that pupils will educate themselves under the influence of especially created situations; and while many adopt these ideas with reservations, especially in the field of arithmetical learning, sufficient approval of these attitudes has been evinced to influence considerably the teaching situation in classrooms in Scottish Primary Schools. Rather generally, progressive School Authorities (districts) are attempting to break down the boundaries of subject delineation and to integrate educational pursuits more widely.

One of the beneficial results of the movement has been the support it has given to those who for many years have been asking for a longer period of preparatory activity in schools before formal studies are attempted, and it is in this respect that its impact is most strongly felt. It is now generally recognised that the child must have built up a store of language based on individual experience before he can attack with understanding the acquisition of mathematical concepts, among others.

The fundamental question which must be answered before any 'programme' or 'anti-programme' can be adopted is, "What purpose do we as teachers expect this study to fulfil for the child?," and to such a query there can of course be no simple answer. Any attempt to particularise what adult skills or specific knowledge must be useful to to-day's children would be rash indeed. Rather must we who are mature think in generalisations, endeavoring to encourage in our pupils the basic qualities of character and mentality which will equip them to face novel situations with confidence. According to progressive educational theories the acquisition of a large amount of factual knowledge is not enough. The end sought, rather, should be to increase the power of logical thinking and to maintain concentration while keeping alive curiosity and imagination. To achieve this kind of result a workable compromise is usually accepted between 'formal' and 'free' approaches, perhaps more readily in Cuisenaire schools than in those using traditional methods in which formality definitely predominates.

To concentrate now upon the teaching of arithmetic in Cuisenaire schools: it is fairly widely recognised, though unevenly put into practice, that children must have time and opportunity in the first instance to acquire a vocabulary of basic size relationships and to comprehend the usefulness of our basic numerical conventions and procedures, and that this informal enlargement of experience should be a natural development in which the child is personally involved. He should be provided with an environment rich with possibilities for exploration in which through the exercise of his curiosity and imagination he can master an understanding of such terms as long, short, light, heavy, few, and many, as well as of positional words like above, below, between, next to, and the like. Achievement of this purpose, in the case of the underprivileged especially, takes time. As a result work with a conventional number system may be postponed for as much as six months to one year after the child's entry into the Primary School at the age of five, so that, although prior to then he has 'counted' to satisfy his play needs and has taken part practically in the as yet unnamed processes of addition, subtraction, multiplication, and division, he will not at this stage have attempted computation.

During this preparatory period the Cuisenaire rods are used frequently and extensively but without number significance, being freely available to children as a kind of concrete apparatus through the use of which they can discover examples of relationships and progressions for themselves while making patterns and constructions of varying complexity.

Stated differently, during the Primary I stage the major part of the school programme is the development of language skills and the creative arts. Arithmetic, except as it occurs in the following of the children's own interests, plays a minor role, with computation limited for the most part to working with numbers to the limit, ten. By Primary II a wide progressive programme of directed instruction combined with free discovery is usually in operation, while by the end of Primary III many teachers expect all but their less able pupils to be able to work quickly and accurately sums (examples) involving the basic processes of addition, subtraction, multiplication, and division with numbers up to 100, as well as having knowledge of simple fractional values enabling them to work such examples as 7/12 x 48, or (3/4 x 16) - (1/3 x 9).

It may be that as much as from forty to sixty minutes per day may be assigned to the mathematical programme, but much less of the time is given over to the working of mechanical sums than was formerly the case, the time saved being devoted to the practical use of simple tables of weight, length, money, time, and capacity and the study of shapes, sets, and graphs on a simple level. Social arithmetic now plays a large part in classroom operations in which the basic numerical facts, first discovered through the use of the Cuisenaire rods, are employed to solve individual problems arising in play or in projects.

It is evident that the Cuisenaire programme tends to lay less emphasis upon abstract computation than does the

"Traditional" programme and, by the same token, far greater emphasis than the latter on the discovery of mathematical relationships and upon their extension through the child's own initiative. By contrast, in most traditional schools the working of sets of mechanical sums occupies a larger proportion of the time allotted to arithmetic.

For this reason the use of paper-and-pencil tests in computation, as in the present inquiry, may have put the Cuisenaire children at a serious disadvantage, in that they were unable to disclose their capabilities in other aspects of arithmetic. For example, the power to reason, one of the characteristics that, it is claimed, the Cuisenaire programme inculcates, was not measured and had virtually no chance to operate and reveal itself. In a word, evaluation when limited to computation, one of the learning outcomes not particularly stressed in this programme, could yield a distorted picture of the true and full accomplishments of children taught according to the Cuisenaire program.[2]

---

[2] Writer's comment. The reader should bear these statements in mind, for they express facts. On the other hand, they do not invalidate the present inquiry. As has been mentioned before, the purpose was not to compare all the results of teaching the Cuisenaire and the Traditional programmes in their entirety. It was, rather, to determine the relative success of the programmes in developing computational skill alone, and in this one respect Cuisenaire children could conceivably prove to be superior to those taught the Traditional programme. In the second place, the principal instrument for measuring achievement in computation, the Common Test, contained no examples which were deemed to be unacceptable as learning outcomes at the end of Primary III in Cuisenaire schools--this in accordance with the judgment of all members of the Cuisenaire panel.

Without in any way drawing invidious comparisons with the Traditional programme, there are perhaps some facts relating to the teaching situation in Cuisenaire schools which should be noted for their possible effect upon the performance of the children. Some proportion of Cuisenaire teachers have had little training in the use of the method, and none of them of course has been taught as a child through this medium,--a condition which could lead to a lack of confidence and skill in handling the materials in the classroom. Then, too, a great variety is observable in the manner in which teachers employ the Cuisenaire apparatus. In some cases children may be allowed free access to the rods and are given adequate opportunity to use them for exploration and discovery. In other classrooms, the rods may be employed only in a limited way, and then for directed lessons on the number facts only. Again, some teachers have the tendency merely to substitute the new length materials for the discrete objects formerly employed, without the disposition to make the required changes in procedure. Still again, in some schools the rods are abandoned when a certain stage of the programme has been reached; in others this is not so. And, last of all, in some schools importance is attached to premature memorisation and the Cuisenaire materials are employed to encourage it, while in others the maxim is honored that through operation comes memorisation, and teachers proceed on these lines. It is therefore possible that in Cuisenaire schools there was more unevenness in the teaching situation than there was in Traditional schools where a familiar medium was used and a more strictly prescribed system of instruction was followed.[3]

As regards the extent of the use of the Cuisenaire programme throughout Scottish schools it can be said that despite the flood of concrete apparatus that has appeared in the last ten years, the Cuisenaire apparatus seems to be holding its own. The lower Primary classes (five to seven years) maintain interest in it in areas in which it has been established for some time, while new Authorities are continually asking for information and, in many cases after instituting study courses, they have adopted it as their concrete medium. It has not, however, except in a few instances been used consistently throughout the Primary Schools of the country (through Primary VI) as was hoped, and it is now being dropped in the majority of schools at the end of the second year. When employed in Primary III and later on, its use very frequently is limited to remedial work or to the demonstration of a new rule. (But see footnote 3.)

Almost without exception teachers who have employed the Cuisenaire programme speak enthusiastically of the enjoyment children derive from its use, the more experienced among them looking upon the materials as an illuminating medium to create interest in, and understanding of the arithmetical studies which children pursue. Perhaps one of the most valuable side products is the fact that most of those who have employed the Cuisenaire apparatus have been forced through their enlarged experiences with them to adopt a more imaginative approach to the subject of arithmetic.

---

[3] Writer's comment. It is to be remembered that in this study all cooperating Cuisenaire schools had used the materials continuously until the end of Primary III, and an effort was made to enlist only Cuisenaire schools in which instruction was thought to be superior in quality.

Computation in the Traditional Programme
(Mr. Robert J. Allan)

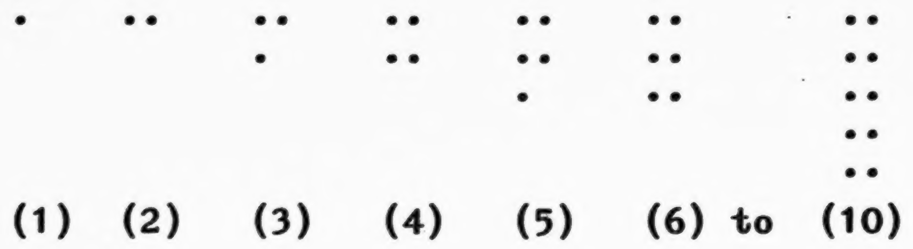
As he scans the statement below, the reader should know that the arithmetic programme described, starting with the paragraph next but one, is no longer in effect in Scottish schools. Rather, the statement is an abbreviated account of the programme taught to children like those tested in this investigation in the early 1960's. Even then mathematics in the Junior School was undergoing change; and in the last year a completely new syllabus has come into force. Emphasis is now placed upon discovery, experiment, and communication; and any order, structure, or pattern in the environment which can be described in terms of numbers or logical relations is viewed as appropriate for mathematical study.

For generations parents have considered the 3 R's the most important part of the Junior School curriculum; and accordingly arithmetic has been accorded a prominent place in the daily schedule. Computation, too, has always been regarded as a critical aspect of the study of arithmetic. At the present moment computational skill has a high priority for children aged 7 to 10; but for children aged 5 and 6 understand ing of elementary uses of individual small numbers and their relation to measuring is of the greatest importance.

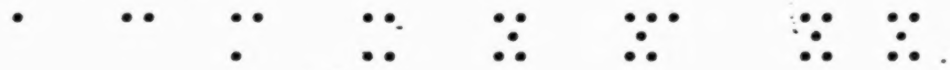
Along with a maze of arithmetical words--bigger, smaller, names and values of coins--the usual starting point for arithmetic in Primary I is counting. Children learn the correct order of number names and learn how to repeat them in one-to-one correspondence with objects enumerated. Every opportunity is taken to get children to see that they can communicate to adults something about groups which interests adults, namely, the numbers of objects in groups (groups of six objects or fewer at the outset).

The use of a particular grouping of objects is stressed as providing the apparatus with which they will study arithmetic. In many schools this takes the form of an arrangement of counters in some geometrical pattern. Two common patterns are:

(a) a system of pairs:



(b) a domino type:



Other patterns (e.g., threes) are occasionally found. Some schools use beads strung on wires, the children originally doing the stringing themselves, but later dealing with permanent sets. Longer wires with as many as twenty and even thirty beads are sometimes found in classrooms.

Children in the earliest stages then are taught
(a) to know the number names and their order, at least to 10;
(b) to use them in enumerating small groups of objects;
(c) to recognise and write the corresponding numerals; and
(d) to recognise and/or create or choose a special pattern of objects (e.g., the domino pattern) for numbers to be used in the study of each number. The composition of each number is examined in ways suitable to the special pattern for that number; thus, five counters are laid out and "patterned," then separated into two groups of two and three, which may or may not be set up in their characteristic patterns. Children then use numerals and signs to record what they have done: e.g., 2 + 3 = 5; 5 - 3 = 2.

Meanwhile counting is extended past 10 and upward to 100. Teachers use a variety of arrangements of materials to give meaning to the sounds "forty-five, forty-six, ... fifty." Children are taught to read and write the new larger numerals with appropriate reference to their apparatus and some idea of the use of zero and of place value in two-digit numbers.

The composition of numbers up to 20 is next studied; e.g., 13 = 8 + 5 = 5 + 8; 13 - 8 = 5; 13 - 5 = 8. Eventually, given any two groups of objects (each fewer than 10), the children should be able to "put them together" either into one group of 10 or less or into a group of 10 plus a remaining group. Children may need some concrete apparatus to demonstrate the "exchange," such as the abacus. Children also may acquire the practice of "counting on"; for example, 3 counted on to 5 gives "six, seven, eight." If the total of the two groups is greater than 10, children will move enough counters or other objects from the smaller group to make the larger group number 10; then the remainder of the small group gives the units figure of the total. This "bridging the ten" in addition, if used mentally as a helpful picture, should be a temporary expedient. "Bridging the 10" in subtraction is sometimes taught as a permanent technique . These additions and the corresponding subtractions should be known by heart, and many applications of them should be provided.

"Counting on" and "bridging the ten" are helps with additions like 28 + 6. The close relationship of this expression to 8 + 6 must be recognised. And the corresponding subtractions (34 - 6 and 34 - 28) are carefully taught. Counters are not suitable for teaching such computations with larger numbers, and something like the "hundred board," or the abacus, or the number line is substituted. Many schools stop using concrete material, for explanations at this stage and pursue more abstract and authoritative methods of teaching additions and subtractions of tens and units, for example, in subtracting by the very popular method of equal additions.

Other related activities are: adding on twos, adding on threes, reciting the even and the odd numbers, counting by twos. Children are taught the value of common coins and solve such simple problems as, "What coins will I use to pay 8d.?", and "What change should I receive from a shilling if I spend 4½ d.?" They are supposed to be able with confidence to measure in inches and feet and have had some experience in weighing and measuring capacity, telling time, and the like. They are expected, too, to be able to use fractions to a limited extent.

The phrase "to a limited extent" needs to be clarified. What is done with fractions depends almost wholly on the textbook employed in a school. On this account there is variation from school to school in the extent and in the details of instruction. In full recognition of this fact, the following paragraphs seem to give a fair summary of practice in Tra. schools.

In Primary II schools generally appear to agree in familiarizing pupils with the fractions ½, ¼, and 1/8; but only as they are used with concrete materials. Thirds and fifths also may be shown, and occasionally sixths; but sevenths and ninths, hardly ever. In Primary III 1/3 and 1/5 are used, if at all, only as equivalent to "short division" by 3 or 5. "Quarters" may be introduced in conversation as preparation for textbook problems in "short division" like "72 apples were shared equally among 4 families. How many did each family get?" "Find 1/5 of 85," a textbook exercise of a sort appearing relatively seldom, is considered as a variant of 85 ÷ 5 or $5\overline{)85}$.

In Primary III the fraction 3/4 (and a few others of the kind in commonest usage) may appear in a textbook problem like "Find the cost of 2 3/4 pounds of steak at 6s. a pound." In such cases the meanings of the terms denominator and numerator are taught. But such instruction is not extended to many fractions of this type; and computational use of 2/3 and 3/5, for example, is postponed at least until Primary IV; for exercises like "Find 3/5 of 55" are viewed as two-step problems of a proportion type.

A very few fractions like 4/4 and 8/8 are presented, but only in connection with actual materials, such as paper and cloth, and then for the purpose of identification rather than for the purpose of computation.

So much for fractions; and to move on now to other topics: The teacher receiving children in Primary III (at the age of 7+) hopes that they will have thoroughly mastered the 100 addition and the 100 subtraction facts. She usually teaches arithmetic with little or no apparatus except with fractions, as noted above, and with such concepts as area and perimeter. The Primary III teacher extends notation and goes on to addition and subtraction of tens and units and also of hundreds, tens, and units. She also begins to deal systematically with multiplication up to "6 times" and to division by divisors up to 6. Each multiplication table is explained as involving the addition of equal numbers, and each division table as implying sharing, at least at first. The learning process is assisted by work with "short multiplications" of the type $\begin{array}{r} 26 \\ \underline{\text{x}\ \ 4} \end{array}$ and with "short divisions" like $4\overline{)76}$.

In general, minimal notice is taken of the relations among operations, and most work is done with formal algorisms. To illustrate: children are told that the division $3\overline{)22}$ means to "share 22 as far as you can among three people," and they are asked to suggest the largest part of 22 which can be shared exactly. However, they are taught to say and to think, "Three into 22 goes seven times and 1 over," as if they were finding the number of 3's in 22 (measurement division). It is expected that the meaning of algorisms will result from recognizing their usefulness in solving problems that involve relevant calculations.

Addition and subtraction of money follow, the same method being used for the subtraction of money as was used for subtracting tens and ones, namely, the method of equal additions.

By the end of Primary III (age 8) the pupil will have worked out many calculations in the addition and subtraction of three-figure numbers, multiplication and division of three-figure numbers by numbers less than 10, addition and subtraction of money and possibly of linear measures. He will also have met numerous "problems" from the school textbook and had to use all four processes to solve them.

In Primary IV the children pursue studies to complete all the multiplication tables up to those with 12 as a factor. "Long multiplication" (e.g., 35 x 436) will be introduced but not "long division" (two- and three-figure divisors). And more attention is paid to problems and formal work with money, weights, and other measures.

One last word: as is implied by the occurrence of such phrases as "in some schools" in the foregoing statement, it would be a mistake to assume that all Tra. schools have at all times been accustomed to teach exactly the same programme in exactly the same manner. On the contrary, there have always been variations both in the aims sought in particular grades and in the methods employed to achieve them. The term "Tra. programme" therefore includes a range of differences in theory and practice. On the other hand, despite dissimilarities, there has been much in common among the Tra. schools, enough so to warrant treating them as a group and as a group quite distinct from that made up of the Cui. schools.

CHAPTER II

DETAILS OF THE PROCEDURE

Subjects

Tests were administered to 1337 Scottish children. This total was materially reduced by two circumstances: (1) a good many children failed to take one or more tests and so, their partial records were discarded; (2) after testing had been started, it was discovered that one large school could qualify neither as a Cui. nor as a Tra. school. Nevertheless as a courtesy the testing was continued despite the valuelessness of the measures obtained.

In the end, complete records were procured from children, of whom 539 were in the Cui. group and were drawn from 18 classrooms in 12 schools, and 570 were in the Tra. group drawn from 18 classrooms in 12 schools. Geographically, the cooperating schools are located in a band approximately 50 miles wide and stretching across Scotland from Aberdeen at the northeast to Ayrshire at the southwest.

The ages of the subjects ranged from 91 mo. to 119 mo., with a mean of approximately 100 mo. for both groups. The testing was done in September, 1966, in the first two weeks of Primary IV. Hence, all subjects had completed three years under a particular program. The last descriptive phrase does not mean that the subjects were equivalent to American children beginning Grade 4, for Scottish children enter school at age five instead of age six.

Ideally, the testing should have been done at a later point in schooling, say, as the Primary VI level, in order to see the relative long-time advantages of the Cui. and of the Tra. programs; but such postponement was not feasible. Beginning with Primary IV, or even with Primary III or Primary II, something very much like the Tra. system of instruction is instituted in most schools which start with the Cui. program. The number of Cui. schools that hold to the Cui. program through Primary IV is very small indeed, and the number of Cui. schools that hold to the program through Primary III is not large. Hence, the testing in this study was done at the last safe point in schooling. Had it been done later than the beginning of Primary IV, there would have been no way of disentangling the effects of the Cui. program which would have been "contaminated" by effects of the Tra. program.

There were several reasons why the investigation was made in Scottish schools. The first is the excellent contacts with the Scottish Council for Research in Education and with Directors of Education (superintendents of school districts, in American parlance) established in the writer's 1962-63 research.[1] These contacts were of inestimable value, for, as on the previous occasion, not a single Director of Education, when approached, refused to take part in the inquiry, and their Head Masters and Head Mistresses were equally cooperative. The second reason is that the Cui. program, by and large, is probably better and certainly more extensively taught in Scottish schools than in American schools. The third reason was one of economy. Scotland being a small country, it was easier to secure a geographic sample of schools than in the United States, without having to send corps of testers on long, expensive trips.

Selection of Schools

Despite what has been said, it was not easy to find schools and classrooms that could be used. The number of Cui. schools that abandon the Cui. program after Primary II greatly exceeds the number that follow that program through Primary III. Not infrequently schools said to be Cui. schools, and tentatively selected, had actually changed to the Tra. program in Primary III and, many of them even in Primary II. By the same token, some reportedly Tra. schools proved, on examination, to have had a year or more of the Cui. program.

---

[1] William A. Brownell, Arithmetical Abstractions: The Movement toward Conceptual Maturity under Differing Systems of Instruction, University of California Publications in Education, Vol. 17, Berkeley and Los Angeles: University of California Press, 1967. 221 pp.

Since it was essential to have "pure" samples of subjects in the Cui. and the Tra. groups, it became necessary to set up two rules:

(1) No Tra. school or class (or said to be such) would be included if at any time the pupils therein had been exposed to the Cui. materials and program, and

(2) No Cui. school or class (or said to be such) would be included unless the Cui. program had been the exclusive basis of instruction in Primary I and II and had been continued through Primary III, at least in the teaching of multiplication, division, and fractions.

The Tests

As stated in Chapter I, four tests were administered to all subjects, three of them in arithmetic computation, the fourth of them an "intelligence" test. Copies of the arithmetic tests will be found in the Appendix.

The "intelligence" test

It seemed unwise to use an American test of "intelligence." Accordingly, on the advice of Dr. A. E. G. Pilliner, then Director of the Godfrey Thomson Unit for Educational Research of the University of Edinburgh, the 8+ Verbal

Reasoning Test was selected. This test was constructed by Scotsmen and was standardized on Scottish children. Since it is not available to the public, no copy can be included in this report.

The 8+ Verbal Reasoning Test contains 65 items, most of them similar to items in American tests, but phrased for Scottish children, and only eight of them requiring the use of simple arithmetic. Special care is taken in introducing the test proper, a period of 15 to 20 minutes being set aside for instructions. The time allowance for the test itself is 35 minutes.

The Unit for Educational Research scored the test papers and standardized the scores especially for the population used in this study. The method used was as follows:

> "The scores as furnished to the writer are given in the standard form of quotients; i.e., the marks gained by the children on the test--the number of questions right--have been converted to a range of scores having a mean of 100 and a conventional distribution. Thus, if a child's score is given as 100, he has shown average ability on the test. Further,
> if a child scores over 125 he is in the top 5% of the children taking the test
> if a child scores over 110 he is in the top 25% of the children taking the test
> if a child scores under 90, he is in the bottom 25%. . . .
> if a child scores under 75 he is in the bottom 5%. . . ."

The arithmetic tests

In the winter of 1965-66 through correspondence with

Scottish leaders in education, steps were taken to set up two three-member panels of arithmetic specialists. The two panel chairmen (Miss Margaret L. Law for the Cui. program; Mr. Robert J. Allan for the Tra. program) each chose two other qualified persons, in each case a Head Master and a classroom teacher. (The teacher on the Tra. panel was also Infant Mistress in her school.)

Identical instructions went to both chairmen. Each with the members of his panel was to prepare a test, disregarding length, which would measure skill in all phases of arithmetic computation taught according to a particular program in the first three school years. The panels worked independently and of course produced quite unlike instruments. Each panel had as many physical meetings as were needed in order to construct the initial form of its test, to try it out in classrooms, and to make revisions. The tests were to be--and were--ready for the writer on his arrival in Edinburgh about the middle of May, 1966.[2]

---

[2] In the winter months, also by correspondence, the writer assured himself that enough schools would cooperate to make the investigation practicable. Dr. D. A. Walker, Director of the Scottish Council for Educational Research, was able to report that all participating schools in the 1962-63 inquiry would be again available. Miss Law was most helpful in assessing the situation in nearby Cui. schools.

The Common (Com.) Test in tentative form was derived by the two panels together in conjunction with the writer, by selecting items from the long original Cui. and Tra. tests. No item was chosen for the Com. Test unless both panels agreed that it represented skills taught in both programs. The resulting test was administered to special samples of Cui. and Tra. subjects and reconstituted several times, until a form was found that was serviceable for 35 minutes of working time and was highly reliable. The final test contains 72 items.

During the tryouts directions to children were steadily improved. Pre-test explanations and practice were provided as needed, using materials printed on the first page of the final test booklet. (1) Addition examples with three or more addends were stated in both the horizontal and the vertical forms, with the instruction, "Add either way." This last expression was made intelligible by studying a sample item. (2) The Cui. children were more accustomed than were the Tra. children with writing missing numbers in places other than as answers; e.g., 12 + ... = 20; ..... - 13 = 13. Since the purpose of such test items was to reveal skill in computation (and not familiarity with ways of expressing examples), it was agreed by both panels to supply explanations and practice for the Tra. subjects. To assist them six sample items printed on the first page of the booklet were worked through prior to the test proper, and extra examples, written on the chalkboard, were employed when necessary.

The Com. Test yielded scores (1) for accuracy (the number of attempts with correct answers), (2) for non-attempts, and (3) for rate of work. To secure measures of the last-named type, children were directed to mark with a large X the item they had just completed when a signal was given at the end of 25 minutes. A count of this item and those preceding it was the score for speed of work.

The special Cui. Test was prepared exclusively by the Cui. panel and was designed to measure skills in computation taught in the Cui. program but not in the Tra. program. The resulting tentative test was tried out with samples of Cui. subjects. It was then modified by the panel in conjunction with the writer, and the new test was administered in mimeographed booklets. Again the test was revised, this time by the panel chairman, one member of the panel, and the writer. Before being released for printing (early June, 1966), the other member of the panel examined and approved it. The final test consists of 60 items, which had proved to be sufficient for a testing time of 35 minutes.

Naturally, no special instructions, prior to the research testing, were necessary for the Cui. subjects except at one point, since, for them, the instrument contained only examples which they were supposed to have learned to compute. This point was to explain to all Cui. subjects the use of brackets []'s in place of the parentheses ()'s which many of them were accustomed to.

The Tra. subjects needed more help. Confronted with an example like [40 + 4] - [10 + 4] = ..., they would not have known what to do, being totally unacquainted with brackets. Again, since the purpose of such items was to measure skill in computation (and not form of expression), both panels approved the idea of supplying explanations and practice with six simple examples of this type (e.g., 4 + [2 x 1] = ...) on the first page of the test booklet before starting the test proper. (Other examples were written on the chalkboard and explained when extra practice seemed to be necessary.) The Tra. children were also reviewed on their ability to compute in examples of the type 2 + ... = 5 and ..... - 4 = 1, with which they would have had previous experience on the Com. Test.

Scores on the Cui. Test were obtained, as for the Com. Test, for accuracy, non-attempts, and rate of work.

The special Tra. Test was constructed in much the same manner as was the Cui. Test. Members of the Tra. panel started with their original long Tra. test, selected items in it which were not suitable to, and were not used in the Com. Test, modified them as seemed desirable, and added enough new items (consistent of course with the Tra. but not the Cui. program) to make a new Tra. test. This test, and later revisions, were tried out with samples of Tra. subjects until a satisfactory final form consisting of 63 items had been developed. Scores like those for the other two arithmetic tests were obtained in the research testing.

In a short pre-test period all subjects, Cui. and Tra. alike, (1) were given more practice with items of the types $3 + \ldots\ldots = 5$ and $\ldots\ldots - 2 = 2$; and (2) were advised that division examples regularly printed in the form $4/\overline{48}$, with answers to be written above the line, could be changed to the form $4/\underline{48}$, for answers to be written below the new line.

Reliability of tests

The reliability of the three final printed arithmetic tests was checked in September, 1966, with samples of children drawn from Cui. and Tra. schools, children who were not otherwise used in the investigation. The test-retest procedure was used in each case, the second testing coming two days after the first. The reliability coefficients were found to be:

Com. Test, 0.93 (N = 234; 125 Cui. subjects from 4 classes in 3 schools; 109 Tra. subjects from 4 classes in 3 schools)

Cui. Test, 0.91 (N = 208; 8 Cui. classes in 5 schools)

Tra. Test, 0.91 (N = 204; 7 Tra. classes in 6 schools)

Information concerning the reliability of the 8+ Verbal Reasoning Test was furnished by Director A. E. G. Pilliner of the Godfrey Thomson Unit for Educational Research, and is as follows:

> "The test retest reliability coefficient was .93. The procedure employed was to administer this test, and a parallel form of it, to the same group of 200 children with an interval between administrations of one week. The group itself was a random sample from a normal school population.
>
> "The coefficient of internal consistency derived by the Ferguson form of the Kuder Richardson formula 20 was .97 . This was based on the data from the test in question administered to the group mentioned above."

The Testing

Each set of four tests was administered by one or another of eleven persons, all girls in the graduating class at Moray House College of Education. These students had been carefully selected by Miss Law and Mr. Allan, both members of the faculty in the same institution. The testers were available full time for three weeks in September, 1966, prior to their starting their course work in the college.

They proved to be not only entirely competent in the testing, but conscientious and responsible as well. In addition to the testing they scored the arithmetic tests and prepared reports to all cooperating Head Masters and Mistresses.

The testers were instructed by the writer on the nature of the study, examined the tests under his supervision, and had at hand always (a) the Complete manual for the 8+ test as well as (b) a mimeographed two-page set of General Instructions for the testing prepared by the writer, and (c) sets of mimeographed instructions for administering the three arithmetic tests.

The order of tests, the first two given one day, morning and afternoon, the last two the day thereafter, was

Cui. subjects	Tra. subjects
(1) 8+ Test of Verbal Reasoning	(1) The 8+ Test of Verbal Reasoning
(2) The Com. Test	(2) The Com. Test
(3) The Cui. Test	(3) The Tra. Test
(4) The Tra. Test	(4) The Cui. Test

It will have been noted that the two samples of subjects differed in order of tests on the second day only, in order to postpone to the end the particular test with which they were relatively unfamiliar.

Organization of Following Chapters

Chapter III, which constitutes Part II of the report, will set forth the results of applying the procedure of analysis of variance to the data for the totals of 539 Cui. and 570 Tra. subjects on the three tests. This part of the report was prepared solely by Dr. Arden K. Ruddell, Professor of Education, University of California at Berkeley.

In the phases of the investigation reported in Part III a different procedure of analysis was employed, as will be explained in Chapter IV. Random samples of 120 subjects were drawn from the original total samples of Cui. and Tra. children, and their test papers were examined in detail. For each of the three tests the content was broken up into sets of relatively homogeneous items, and the success of the smaller samples was studied both on each set and on each separate test item.

Chapter V is devoted to a comparison of the records of the sub-samples (hereafter referred to simply as samples) with respect to the Com. Test, an achievement test for both the Cui. and the Tra. subjects.

Chapters VI and VII are devoted to the results of attempts to identify extent of transfer of learning, first on the part of Cui. subjects on the Tra. Test, then on the part of Tra. subjects on the Cui. Test.

Part IV contains but a single chapter, in which relevant research will be reviewed and the results of the present inquiry will be summarized.

PART II

AN EXAMINATION OF THE RESULTS BY ANALYSIS OF VARIANCE

CHAPTER III

MAIN EFFECTS AND INTERACTIONS OF VARIABLES IN THIS INVESTIGATION

Introduction

It must be obvious to the informed reader that a variety of analyses are available to the investigator. These options were considered preliminary to setting up the study design and the gathering of data. Appropriate precautions were exercised in selecting the samples to be utilized in order to fulfill the assumptions underlying the statistical tools to be employed in the data analysis described in this part of the report. All sub-sample analyses are based on a random selection of each sub-sample from all possible members of the total sample who meet the sub-sample criteria. The analysis of variance program utilized in this study required that an equal number of cases be placed in each cell. This requirement necessitated a random selection of cases for each cell. In reporting differences between mean scores of groups of pupils only those differences exceeding the .01 level of probability are considered to be statistically significant. All differences falling between the .01 and .05

levels of probability are called to the attention of the reader as worthy of further exploration.

Total Group Results--A First Impression

Relevant information about each child which was available at the outset of the study is summarized in Table 1. A measure of intelligence, the chronological age, and the total time devoted to the study of mathematics were the three variables studied in order to determine the similarity of the two groups of subjects selected for this investigation. A quick inspection of the information in Table 1 reveals that no important differences existed between the Cui. and the Tra. children with respect to these three variables. Tests of the differences between the mean scores and the homogeniety of the variances support this observation.

Table 1

Intelligence, Time, and Chronological Age
for the Total Cuisenaire and Traditional Samples

		Measure of Intelligence	C.A. Months	Time	Number of Schools
Cui.	Mean	99.37	99.98	653.98	
Sample	S.D.	14.81	4.89	151.23	18
N=539					
	Range	68-142	91-119	475-960	
Tra.	Mean	99.07	99.40	678.24	
Sample	S.D.	14.70	4.77	175.98	18
N=570					
	Range	68-140	91-117	375-955	

There is no question as to the comparability of the two groups of subjects with respect to chronological age and the measure of intelligence used by the Scottish schools. We can assume with a high degree of confidence that the two samples were drawn from the same population. Consequently, an analysis of variance was deemed appropriate for testing results of the post-tests.

Further explanation of the time factor may be appropriate at this point. The headmaster or headmistress of each school was asked to estimate the number of minutes each week devoted to mathematics instruction at each grade level. Since the pupils in this investigation had completed three full years of formal schooling, the data reported in Table 1 represent a sum total of the number of minutes per week estimated for the first three years of instruction. (That is, a total of 675 minutes could represent a first year of 200 minutes per week, a second year of 220 minutes, and a third year of 255 minutes.) For the reader's convenience these totals will be reduced to the average number of minutes of instruction time per day; thus, a 675 total would represent an average of 45 minutes of instruction time per day during the first three years of schooling.

To refer to Table 1 once more, the mean time of instruction in the two groups is comparable. An average of approximately 44 minutes per day was devoted to mathematics instruction in the Cui. schools while the Tra. schools spent an average of about 45 minutes per day. It should be observed that the lower end of the range for the Tra. schools dips somewhat below the lower end of the range for the Cui. schools. This difference actually represents an average of approximately six minutes of instruction time per day during the first three years of school. It may be a relevant factor in one of the more detailed analyses to be reported later in the chapter.

The data recorded in Table 1 constitute adequate evidence for assuming comparability of the two groups of subjects in terms of three relevant variables. Attention is now directed to the results of each total sample on the three criterion measures reported in Table 2. Each group of subjects was administered three arithmetic tests; namely, the Com. Test, the Cui. Test, and the Tra. Test described in Chapter II. The mean score for each sample on each test is reported in Table 2 as well as the mean number of attempts on each test.

In general, the data summarized in Table 2 reveal exactly what one would anticipate. The Cui. children scored significantly higher than the Tra. children on the Cui. Test, while the Tra. children scored significantly higher than the Cui. children on the Tra. Test. On the Com. Test there is no significant difference between the two groups as determined through the use of a t-test of difference between means.

Table 2

Mean Achievement Scores and Mean Attempts by the Total Cuisenaire and Traditional Samples

		Achievement			Attempts		
		Com. Test	Cui. Test	Tra. Test	Com. Test	Cui. Test	Tra. Test
Cui. Sample N=539	Mean	39.10	21.78	26.69	56.55	45.21	45.20
	S.D.	15.85	13.65	15.15	14.47	14.56	14.24
	Range	0-70	0-60	0-60	14-72	1-60	2-63
Tra. Sample N=570	Mean	37.37	15.43	30.83	54.74	40.21	46.92
	S.D.	15.59	8.73	15.06	14.16	16.01	12.75
	Range	0-69	0-60	0-60	6-72	4-60	9-63

The relationship between the number of attempts on each test and the accuracy scores is deserving of comment. The ratio between the number of items attempted and the number of items correct is summarized in Table 3.

Table 3

Percent of Attempts Completed Correctly

	Common Test	Cui. Test	Tra. Test
Cui. Sample	69%	48%	59%
Tra. Sample	68%	38%	66%

Obviously, the Cui. test was more difficult for both groups of subjects than either of the other tests. No intelligent comparison of accuracy scores by one group on two or more of the tests is possible since the relative difficulty of the items on the three tests differs greatly.

In summary, the total group results imply that no important differences between the two groups of children can be detected. Both groups performed at the same level in attempts and in accuracy scores on the Common Test. Each group of children scored significantly higher on the test designed for them. The test designed for the Cui. program proved to be more difficult for each group of subjects than the test designed for the Tra. program.

Sub-Sample Performances--A Second Impression

More often than may be suspected gross results in an investigation obscure relevant information about an important sub-sample and do not reveal significant interactions of two or more variables. Not only should the two treatment groups be considered as a source of variation, but the intelligence level of the children, the time devoted to the study of mathematics, and the mathematics achievement level of the subjects should be taken into account. It is entirely possible that a more detailed examination of these sources of variation or interactions between them will reveal valuable information otherwise obscured.

Sub-samples were arbitrarily determined by utilizing the top 40 percent and the bottom 40 percent of the subjects' scores in each treatment group on each of the variables; resulting in a high and low sample in intelligence, a high and low sample in mathematics achievement, and a long and short time sample in studying mathematics. The middle 20 percent of each range of scores was deliberately eliminated from the analyses in order to minimize the overlapping of the two extremes on each variable.

The comparability of the sub-samples in intelligence and length of time devoted to studying mathematics is summarized in Table 4. (The use of mathematics achievement as a condition for subject classification is discussed in a later section.)

The simultaneous analysis of two variables is relatively simple; whereas, an analysis of four variables simultaneously is considerably more complex because of the difficulty in obtaining large enough sub-samples based on a four-way classification. For example, in this study it was easy to obtain sub-samples classified by treatment group and intelligence. There are an adequate number of high intelligence and low intelligence subjects in each treatment group to warrant an analysis. However, it was virtually impossible

Table 4

Mean Scores of High and Low Sub-samples on Intelligence and Time

		Intelligence		Time	
		High	Low	Short	Long
		N=193	N=212		
Cui.	Mean	115.31	84.84	495 min.	805 min.
Group	S.D.	8.15	6.95	(33 min/day)	(52 min/day)
	Range	105-142	68-94	475-510	710-960
		N=234	N=219		
Tra.	Mean	112.95	84.20	445 min.	840 min.
Group	S.D.	7.92	8.10	(30 min/day)	(56 min/day)
	Range	105-140	68-94	375-525	725-955

to obtain an adequate sub-sample based on all four sources of variation. Only three subjects could be identified in the Cui. group: high intelligence; low in arithmetic achievement; and, studying for a long period of time. On the other hand, a relatively large sub-sample (87 subjects) was classified as Tra. group: high intelligence; high achievement; and long time in study. In order to obtain an adequate number of cases in each cell for the analysis of variance, no more than three variables were ever used in any given analysis.

After all pupils were classified according to the three sources of variation under consideration, the cell with the least number of subjects served as the basis for determining the number of cases to remain in each cell for analysis. Scores were randomly selected out of each cell until the number of cases in the cell equalled that of the criterion cell with the least number of cases.

A three-way classification of subjects on the basis of treatment group, intelligence, and time provided the conditions for an examination of accuracy scores and the number of attempts (Att.) on each of the three tests. These data are summarized in Table 5a.

Table 5a.

Mean Accuracy Scores and Attempts for Each Sub-Sample on All Tests

				Cui. Group		Tra. Group	
				Time		Time	
		Test		Short	Long	Short	Long
Intelligence	High	Com.	Att.	65.02	65.79	60.00	69.02
			Ac.	47.94	55.40	49.46	56.34
		Cui.	Att.	46.61	54.34	40.00	49.58
			Ac.	26.57	40.31	21.09	25.40
		Tra.	Att.	49.87	49.58	51.00	58.95
			Ac.	33.89	39.00	41.26	50.14
	Low	Com.	Att.	42.63	52.66	47.79	49.87
			Ac.	21.11	30.94	20.80	22.37
		Cui.	Att.	34.63	44.18	47.37	39.37
			Ac.	8.77	14.66	7.91	8.60
		Tra.	Att.	32.92	45.97	42.13	44.34
			Ac.	10.89	20.43	16.66	21.31

N = 38/cell = 304 total

Main variables--treatment, time, intelligence.

An inspection of Table 5a indicates that pupils in all categories attempted (Att.) many more items on each test than were completed correctly (Ac.). Of course this is to be expected. A more salient result is the fact that in every instance, the subjects who studied for a long period of time obtained a higher accuracy score than subjects who studied for a short period of time. Perhaps the most notable observation rests in the fact that the High Intelligence Tra. subjects obtained higher accuracy scores on the Common Test than the High Intelligence Cui. subjects, while the Low Intelligence Cui. subjects scored higher than the Low Intelligence Tra. subjects on the same test. (The reader will recall from the data presented in Table 2 that no notable differences were discernable between the Cui. sample and the Tra. sample as determined by gross mean scores on the Common Test.) This observation suggests the possibility of an interaction between treatment as a source of variation and the intelligence of the subjects in the study.

In order to confirm (or reject) the initial observations of raw data, the accuracy scores summarized in Table 5a were treated by an analysis of variance resulting in the F-values reported in Table 5b. As one would normally expect, significant differences between mean scores occurred when High and Low Intelligence groups were compared. Significant differences were obtained also, when Short and Long periods of instruction time were compared. These differences were independent of the treatment sample in which the subject resided thus supporting the general contention that subjects of high intelligence will achieve at a higher level than subjects with a low intelligence. Furthermore, they give credence to the belief that pupils will attain higher levels of achievement if subjected to longer periods of instruction.

Table 5b

Analysis of Variance Resume
For Accuracy Scores

	F - Value		
Source of Variation	Com. Test	Cui. Test	Tra. Test
1. Between Cui. and Tra. Groups	1.70	58.14*	21.73*
2. Between Long and Short Time	28.28*	47.27*	27.28*
3. Between High and Low Intelligence	553.75*	420.17*	309.57*
Interaction Between 1 and 2	3.33	16.68*	0.04
Interaction Between 1 and 3	5.49	14.17*	4.82
Interaction Between 2 and 3	0.37	10.28*	0.01
Interaction Between 1,2 and 3	2.52	1.39	2.57

*F - Value of 6.66 = .01 level of probability
3.86 = .05 level of probability

The F-value on the Com. Test indicates no important difference between the two treatment groups. This statistic supports the earlier conclusion when the gross mean scores of the two groups were submitted to a t-test. Thus, random sampling does not cause a change in previously reported results. Further confirmation of earlier analyses is found in the F-value for the two treatment groups on the Cui. Test and the Tra. Test. The Cui. children scored significantly higher than the Tra. children on the Cui. Test, while the Tra. subjects scored significantly higher than the Cui. subjects on the Tra. test.

Interactions between sources of variation resulted in three significant F-values on the Cui. Test and two additional F-values at a level deserving of special comment (F-values for interactions between 1 and 3 on the Com. Test and the Tra. Test). Most likely, the interactions are the result of the unusually high accuracy scores of the Cui. Sample classified as Low Intelligence and Long in instructional time. As compared with all other sub-samples classified as Low Intelligence, this particular sub-sample scored considerably higher than would be expected. Especially noteworthy is the mean accuracy score on the Tra. Test of the Low Intelligence, Long Instructional Time, Cuisenaire sub-sample. This group of subjects scored higher on the Tra. Test than one sub-sample in the Tra. group, which is the only instance of a sub-sample of one treatment group obtaining a higher score on a test designed for the other treatment group. These same children scored significantly higher on the Com. Test than any other sub-sample classified as Low Intelligence.

The F-values reported in Table 5c provide a basis for interpreting the data in Table 5a regarding the number of items attempted on each test by the subjects in each sub-sample. Generally consistent with all previously determined results, subjects classified as High Intelligence attempted a significantly greater number of items than subjects classified as Low Intelligence. Also, children subjected to a Long instructional period scored significantly higher than children participating in a short time of instruction.

The most notable result in the analysis of Attempts is the lack of a significant difference between the two groups on the Cui. Test. Generally, the Tra. subjects attempted just as many items on the Cui. Test as the Cui. subjects. Perhaps the unusually high number of attempts by the Low Intelligence, Short Time, Tra. sub-sample accounts for the lack of an overall difference between the two groups. Most certainly, this sub-sample, as well as the Low Intelligence, Long Time, Cui. sub-sample account for the significant interactions which occurred.

Table 5c

Analysis of Variance Resume
for Attempts

Source of Variation	F - Value Com. Test	 Cui. Test	 Tra. Test
1. Between Cui. and Tra. Groups	0.01	0.15	10.26*
2. Between Short and Long Time	15.19*	6.44	16.48*
3. Between High and Low Intelligence	141.78*	10.29*	60.82*
Interaction Between 1 and 2	0.55	6.73*	0.27
Interaction Between 1 and 3	0.01	3.97	0.21
Interaction Between 2 and 3	0.17	4.00	1.81
Interaction Between 1, 2 and 3	8.33*	6.13	11.42*

*F-value of 6.66 = .01 level of probability
3.86 = .05 level of probability

Arithmetic achievement--a main variable.

Achievement in a subject matter area is frequently used as a criterion for pupil classification in conducting a two- or three-way analysis involving other critical variables in an instructional program. Obviously, a significant difference between high and low achievement groups would be expected on any measure of the subject matter discipline under question. However, important interactions may occur when subject matter achievement is considered simultaneously with length of instructional time, intelligence, and/or type of instructional plan.

No pre-investigation measure of achievement in mathematics was available to the investigator in this study. Therefore, pupil scores on the mathematics test designed for the instructional program to which they were assigned provided the basis for classifying subjects by achievement. Cui. subjects were assigned to high and low achievement categories as a result of their scores on the Cui. Test and Tra. subjects were assigned to the high and low achievement categories as a result of their scores on the Tra. Test. This classification technique assumes that the two tests are equally discriminating as measures of achievement in mathematics.

Since the Cui. and Tra. test results were employed in classifying pupils by achievement, only mean accuracy scores on the Com. Test were submitted to examination by an analysis of variance. Results on the Com. Test under a three-way classification scheme involving time, treatment, and achievement as sources of variation are summarized in Tables 6a and 6b.

The obvious differences in mean scores (Table 6a) between the High and Low mathematics Achievement groups is confirmed by the F-values reported in Table 6b. The most important result in this analysis emanates from the lack of a difference between subjects in the Short and Long instructional time categories. Overall, there is no significant

Table 6a

Mean Accuracy Scores on the Common Test With Achievement and Time as Sources of Variation

Mathematics Achievement	Treatment Group	Time Short	 Long
High	Cui.	53.02	53.64
	Tra.	49.74	49.98
Low	Cui.	21.64	26.87
	Tra.	20.61	24.91

N = 46/cell = 368 total

Table 6b

Analysis of Variance Resume for Accuracy Scores with Achievement and Time as Sources of Variation

Source of Variation	F-Value Cui. Sample	 Tra. Sample
1. Between High and Low Achievement	373.25*	335.16*
2. Between Short and Long Time	3.78	2.36
Interaction between 1 and 2	2.35	1.89

*F-Value of 6.80 = .01 level of probability
3.86 = .05 level of probability

difference between the two groups of children as a result of time devoted to studying mathematics. When they are further classified by mathematics achievement and treatment, there appears to be a significant difference between the mean scores of the Long and Short Time subjects classified as Low in Mathematics Achievement in the Cui. Treatment group. This difference proved to be significant at the .01 level of probability through the use of the t-test. Thus, a significant difference between two sub-samples is identified when a difference between mean scores of the total samples did not occur.

By inspection of the data presented in Table 6a one is impressed that the Cui. sub-samples consistently scored higher than their Tra. counterparts on the Com. Test. The differences between treatment groups when classified into High and Low achievement categories were analyzed by use of a t-test, none of which was found to be significant at the .01 level of probability.

The last analysis completed for this part of the investigation was based on a three-way classification in which treatment group, mathematics achievement, and intelligence were the main sources of variation. These data are summarized in Tables 7a and 7b. Normal expectations were fulfilled with respect to achievement and intelligence levels (i.e. the High Achievement subjects scored significantly higher than the Low Achievement subjects and the High Intelligence sub-sample scored significantly higher than the Low Intelligence sub-sample.) It may be noteworthy that the differential between the High and Low Achievement sub-samples is considerably greater in the Cui. group than in the Tra. group.

Table 7a

Mean Accuracy Scores on the Common Test
With Achievement and Intelligence as Variables

			Intelligence	
			High	Low
	High	Cui.	53.28	48.56
Achievement		Tra.	49.00	39.53
	Low	Cui.	33.28	16.00
		Tra.	34.63	21.74

N = 19/cell = 152 total

For the total sample, no significant difference was obtained between the two treatment groups (grand mean for Cui. subjects = 37.78; for Tra. subjects = 36.22). This result is consistent with all previous analyses regarding the accuracy scores of the total treatment groups on the Com. Test. However, some interesting interactions were identified as a result of this three way classification of subjects.

Table 7b

Analysis of Variance Resume for Accuracy Scores With Achievement and Intelligence as Variables

Source of Variation	F-Value Cui. Sample	 Tra. Sample
1. Between High and Low Achievement	119.79*	63.40*
2. Between High and Low Intelligence	20.99*	30.67*
Interaction Between 1 and 2	6.84	0.72

*F-Value of 6.99 = .01 level of probability
3.86 = .05 level of probability

If intelligence and achievement each had an equal impact on Com. Test accuracy scores, and if there is no other basic factor underlying Achievement and Intelligence to account for the difference, then the mean scores of the upper right cell (High Achievement, Low Intelligence) should approximate the mean scores in the lower left cell (Low Achievement, High Intelligence). The F-value for the Tra. sample indicates no interaction between achievement and intelligence; whereas, the Cui. sample F-value denotes the strong possibility of an interaction between these two variables. For the Cui. group, mathematical achievement seems to have played a more important role than intelligence as the subjects respond to the items on the Com. Test.

It has been determined that no significant difference occurred on the Com. Test between the total samples of the two treatment groups. It should be noted, however, that the High Achievement, Low Intelligence, Cui. sub-sample scored significantly higher than the Tra. sub-sample counterpart. For the Low Achievement, Low Intelligence sub-samples it was the Tra. treatment group that scored higher. These differences would suggest that subjects classified as High Achievers tend to profit more from the Cui. program while subjects classified as Low Achievers tend to learn more in a Tra. program.

Relationships between test scoring.

The correlation matrix presented in Table 8 indicates the relationships of the various measures obtained on the subjects in the two treatment groups of this investigation.

Particular attention is called to the two highest correlations obtained. Scores of the Tra. group on the Tra. Test and the Com. Test produced a correlation of .86 and the scores of the Cui. group on the Cui. Test and the Com. Test resulted in a correlation of .85. These two correlations indicate that subjects who scored high on the test designed for their program would generally score high on the test consisting of items common to both programs. Subjects scoring

Table 8

Correlations Between Mathematics Tests Scores, Chronological Age, and Intelligence Test Score

	Cui. Test	Tra. Test	C.A.	Intelligence	Sample
Com.	.84	.86	.12	.73	Tra.
Test	.85	.76	.13	.67	Cui.
	Cui.	.76	.07	.65	Tra.
	Test	.59	.07	.63	Cui.
		Tra.	.15	.64	Tra.
			.19	.59	Cui.
			C.A.	.06	Tra.
				-.17	Cui.

low on one test would most generally score low on the other test. The correlation between scores on the mathematics achievement test and the measure of intelligence merits special comment. Scores by the Tra. subjects on the Tra. Test and the Intelligence Test correlated .64, while the Cui. Group scores on the Cui. Test and the Intelligence Test produced a correlation of .65. Correlations of this magnitude do not indicate a particularly high relationship between mathematics achievement and intelligence as measured in this investigation. Results emanating from these two main effects produced some conflicting conclusions, thereby leaving the reader to his own preference for final interpretation. When intelligence is the basis for classifying subjects, time makes a difference; when achievement is the variable for classification, time is not important.

Concluding Statement

Conclusions

A relatively high standard was established for determining statistically significant differences between samples. No difference was considered to be significant if it could be attributed to chance alone more than one time out of a hundred. An analysis of variance technique was employed to examine the main effects of the variables (treatment, time, intelligence, mathematics achievement) and their interactions. In a few instances a t-test was employed to determine the difference between two sub-samples.

For the most part, analyses based on the Common Test scores will prove to be of greater relevance than results on other tests. When deemed appropriate and pertinent, analyses of results on the Cui. Test and the Tra. Test have been reported. The evidence amassed in this section of the report serves as a base for these conclusions:

a. with respect to the primary sources of variation.

1. Treatment group. No important difference between the mean scores of the total Cui. sample and the total Tra. sample occurred on the Com. Test.

   As would be expected, the Cui. sample scored higher than the Tra. sample on the Cui. Test.

Likewise, the Tra. subjects scored higher than the Cui. subjects on the Tra. Test.

2. Intelligence. In every instance involving a comparison of subjects classified as High or Low in intelligence, the High group scored significantly higher than the Low group.
3. Achievement. Subjects classified as High and Low in mathematics achievement were compared in several analyses with the High group always scoring significantly higher.
4. Time. In the three-way analysis in which pupils were classified according to treatment, intelligence and time, the subjects submitted to a Long period of instruction scored significantly higher than the subjects participating in a Short period of instructional time.

   When the three-way analysis was based on treatment, achievement, and time as classification variables, no significant difference between the scores of the subjects in the two lengths of instructional time occurred.

   This particular result suggests that if mathematics achievement is the primary basis for grouping children for instruction under either instructional program, then the length of the instructional period will not be a crucial factor in the accomplishment of the subjects. If, on the other hand, intelligence is the factor for grouping children, then the length of instructional time will have an effect on success in either program.

b. with respect to interactions of the main variables.

1. A significant interaction between Treatment Group and Time is accounted for by the unusually high mean scores of the Cui. subjects subjected to a Long Period of Study.
2. All interactions between Treatment and Intelligence are probably attributed to the scores of the Low Intelligence subjects in the Cui. Treatment group. These scores are higher than one might normally expect.
3. There was no interaction obtained between Treatment and Achievement level of the subjects.
4. The interaction between Intelligence and Time probably stems from the fact that Intelligence plays a more powerful role in student success on the mathematics tests than length of study time. Consequently, High Intelligence, Short Time sub-samples scored considerably higher than Low Intelligence, Long Time sub-samples.

5. In contrast to the immediately preceeding finding, there was no interaction between Achievement level and Time.
6. An interaction between Achievement and Intelligence may be accounted for by the differential between the High Achievement, Low Intelligence sub-sample and the Low Achievement, High Intelligence sub-sample; the former scoring considerably higher than the latter. This differential suggests that Achievement (as measured in this study) has a greater effect than intelligence in so far as success on a mathematics test is concerned.

Recommendations

A statistically significant difference between the mean scores of two samples does not necessarily justify major curriculum revision. Each recommendation emanating from the data presented in this chapter is dependent upon a spread of not less than 10% of the total number of items on the test under analysis. This arbitrary decision on the part of the writers is deemed necessary in order to make a recommendation that is educationally sound and worthy of serious curriculum modification. (See pages 75-76 for further explanation of the rationale underlying this procedure.)

1. The mean scores of the two treatment groups on the Common Test will not permit a clear-cut recommendation of the superiority of one instructional program over the other. Any major differences obtained were between two sub-samples and most frequently were the result of an interaction between two main effects.

2. If the Cui. program is to be adopted then a Long period of time should be utilized for instruction. The Cui. subjects in both the High and Low Intelligence categories scored higher on the Com. Test if they were in a Long period of instructional time. Differences of an adequate magnitude did not occur in the Tra. sub-samples. This result is all the more noteworthy since the differential between the Short and Long Time periods was not as great for the Cui. sample as for the Tra. sample (see Table 4). Perhaps it should be pointed out that the Low Intelligence Cui. sub-sample gained greater benefit from a Long instructional period than any other sub-sample.

3. The evidence presented in Table 6a will not permit any recommendation regarding the value of a long instruction period when mathematics achievement is a factor in classifying subjects. The only difference obtained that is worthy of comment concerns the Low Achievement, Cui. subjects in a Long time period. This sample scored higher than its counterpart in the Short time period which lends

support to the suggestion that children identified as low in ability (whether by intelligence or by achievement) will profit more from the Cui. program if placed in a long period of instruction time.

4. Any major benefits derived from the Cui. program are further substantiated by the data in Table 7a. When the subjects were classified by both achievement and intelligence, a major difference between the Cui. and Tra. subsamples was obtained in the cell classified as High Achievement, Low Intelligence. The Cui. Program appears to be especially effective for children falling toward the lower end of the intellectual scale but who are achieving toward the upper end of the mathematics scale. The Cui. Program seems to be of least value for children classified as Low in both Intelligence and Achievement.

PART III

RELATIVE ACHIEVEMENT AND RELATIVE EXTENT OF TRANSFER, ACCORDING TO TYPE OF SKILL

CHAPTER IV

PURPOSE AND PROCEDURE

The purpose of Part III is to report the results of comparing the records of Cui. and Tra. subjects with respect to kinds of computational skill; for example, addition of whole numbers. (By contrast in Part II types of skill were ignored, and comparisons were based on scores on tests as a whole.) Attention will be given (a) to the relative achievement of Cui. and Tra. subjects on skills taught to both groups (the Com. Test),--this in Chapter V; but, more especially, (b) to evidence of transfer of learning,--the Cui. subjects on the Tra. Test, and the Tra. subjects on the Cui. Test,--this in Chapters VI and VII.

Program Samples

Through use of a table of random numbers a sample of 120 subjects was drawn from the original total group in each program. These samples represent 22 percent of the original Cui. group and 21 percent of the original Tra. group.

According to the data in Table 9 the samples of 120

Table 9

Comparability of Total Groups and Samples,
General Factors

Categories of Subjects	C.A. in months		Instructional Time, 3 yr.		Intelligence Test Scores	
	Cui.	Tra.	Cui.	Tra.	Cui.	Tra.
Total groups						
Mean	99.98	99.40	44 min.*	45 min.*	99.37	99.07
S.D.	4.89	4.77	--	--	14.81	14.07
Samples						
Mean	99.30	99.25	44 min.*	45 min.*	99.07	100.21
S.D.	4.65	4.44	--	--	14.55	15.03

No difference between means is significant at the 0.01 level. Only one, that for C.A., is significant at the 0.05 level.

*Total time in the three-year period reduced to average number of minutes per day.

subjects resembled closely their parent populations in age and in intelligence test scores, and were nearly identical in the total amounts of instructional time[1] they had been accorded.

---

[1] See pp. 43-45 for the method of determining instructional time.

Table 10 reports comparative data for the three arithmetic tests. In the case of the Cui. subjects differences between accuracy means of the reduced sample and of the original sample are 1.18, 0.12, and 0.47 on the Com., the Cui., and the Tra. Tests, respectively, and the differences between the two means for rate of work are all 1.21 or less. For some reason the reduced Tra. sample has larger accuracy means on all three tests than has the total Tra. sample,--1.46, 1.42, and 1.43--as well as two larger means for rate of work,--0.59 and 0.38 (Com. and Cui. Tests). None of the differences mentioned is reliable.

Method of Analysis

1. Total scores on sets of items.

The content of the three arithmetic tests was broken into sets of comparatively homogeneous items. Total scores for accuracy (Ac.) and for non-attempts (NAs) on each set were obtained and are reported in a series of separate tables for each test. Tables 11 to 17 for the Com. Test in the next chapter are illustrative. Means for the two samples on each set of items are presented, along with S.D.'s, the latter however only when the number of items in sets is seven or more.

Table 10

Comparability of Total Groups and Samples, Scores on Arithmetic Tests

Instruments, measures, and groups	Cui. subjects		Tra. subjects	
	Mean	S.D.	Mean	S.D.
Com. Test				
Accuracy				
Total groups	39.10	15.85	37.37	15.59
Samples	40.28	16.10	38.83	17.15
Rate				
Total groups	56.55	14.47	54.74	14.16
Samples	56.69	14.69	55.19	14.74
Cui. Test				
Accuracy				
Total groups	21.78	13.65	15.43	8.73
Samples	21.90	13.59	16.85	9.35
Rate				
Total groups	45.21	14.56	40.21	16.10
Samples	44.57	15.77	41.52	15.48
Tra. Test				
Accuracy				
Total groups	26.69	15.15	30.83	15.06
Samples	26.22	16.75	32.26	15.72
Rate				
Total groups	45.20	14.24	46.92	12.75
Samples	45.83	13.84	46.77	13.13

The following differences between means are significant at the 0.01 level: for the total Cui. group over the total Tra. group in both accuracy and rate on the Cui. Test; for the total Tra. group over the total Cui. group in accuracy on the Tra. Test.

No differences between pairs of total groups and of samples within the same program approaches significance.

Three differences are significant at the 0.05 level: total Cui. group vs. total Tra. group for accuracy and rate on the Com. Test; total Cui. group vs. total Tra. group for rate on the Tra. Test.

2. Scores on specific items.

For each item in a given set three measures were recorded for each child, these measures being (a) for correct answers, (b) for errors, and (c) for NAs (non-attempts). The resulting tabulations will not be reproduced in this report, but citations to the data therein will occur frequently.

One procedure employed, being based upon negative measures (lack of success,--errors and NAs), rather than on positive measures, such as correct answers, could lead to misinterpretation, for a possible effect is to inflate the significance of the former. To illustrate: on a given item the Cui. sample could make eight errors and omit seven more, making a total of 15 negative responses. The record of the Tra. sample on the same item might be nine errors and 11 NAs, a total of 20 negative responses. The difference between the totals of negative responses, 5, may seem significantly large. Yet, it need not be because of the fact that each sample had 120 chances on the item. The totals of accuracy scores (Cui., 105; Tra., 100), when divided by 120, yield accuracy percentages of 87.5 and 83.3, the difference between the two being merely 4.2 percentage points.

3. Non-attempts (NAs)

Correct and erroneous answers are readily identified: a subject attempted a computation and was successful or unsuccessful. There is no ambiguity in interpreting responses of this kind. Such, however, is not always the case with NAs. For the purpose of studying relative performances of the samples on sets of items as whole, lack of an answer on a given item was regarded as a non-attempt and was so counted in the tables reported, such as Tables 11 to 17 for the Com. Test.

In the analysis of item-by-item performances, on the other hand, a somewhat different practice was adopted. When a subject left blank the answer space for a given item, the instance was always recorded as an NA, but the interpretation varied. If he had made no attempt on items following that in question, his failure on the given item could have been caused by lack of time, and note was made of that fact. On the other hand, if he skipped the item in question, as he was allowed to do, but attempted even one item beyond that point, his NA was very probably the result of a choice: he feared that he could not make the computation called for.

4. Reliability and significance.

a. Scores on tests as wholes. Assume that on a 40-minute 50-item test the mean scores for accuracy of two large comparable hypothetical groups taught according to unlike programs of instruction are 37.5 and 39.1. Suppose further that the difference between the means, 1.6, is highly reliable according to all statistical criteria. From this fact it may be inferred that the program taught the group with the larger mean is superior to that taught to the other group,--and, more than that, that it should be adopted in schools, transplanting the other program when it is in effect.

Yet, in the situation described the difference between means, 1.6, represents a difference of 3.2 percent in performance on the test as a whole. In educational practice changes in programs are hardly justified when based on such slight statistical evidence. And educational practice in such circumstances is sound, especially when it is recognized that the evaluative evidence (scores on a single test) is so limited. Left out of consideration are data on many types and unlike amounts of learning not represented in the one test.

In a word, a difference between means, may be statistically reliable but educationally meaningless. This statement is not to be construed as an attack on statistics, but rather is intended to stress the effect of applying quite different standards in interpreting the significance of the same statistical measure. If a difference between means is highly reliable (a statistical term), then one knows that this difference is not a matter of chance, but can be confidently expected to recur whenever the given situation recurs. But if this difference amounts to nothing of educational importance, then that difference can be safely ignored in determining educational practice.

In Chapter III the basic data were the total scores of two large groups of subjects on three arithmetic tests. Moreover, the tests are known to have reliability coefficients of not less than 0.91. Differences between means could therefore be checked statistically for reliability. Nevertheless, reliable differences had still to be interpreted with regard to their educational significance. The standard employed in this latter check, admittedly arbitrary, was that differences had to represent 10 per cent of the highest possible score. Thus, on the 70-item Com. Test, a difference in accuracy had to amount to 7.0 points to be taken seriously.

b. Scores on sets and groups of items. In Part III the basic data are scores of two randomly selected samples of 120 subjects on sets or groups of similar computational items. In only a few instances in the three tests are there

10 or more items in a set or group. In the Com. Test, for example, one set comprises 20 items with sub-groups (or groups, as they will be called) of 12 and eight items. Other sets contain 15 items (groups of five, four, and six), five items, seven items, six it ems (groups of three and three), six items and 13 items. While no time limits were imposed on sets, it is improbable that the average subject spent more than ten minutes or so on any one. Hence, the scores on sets--to say nothing of scores on groups within sets and on individual items--militate against the applicability of customary procedures of testing reliability between pairs of mean scores.

In Chapters V - VII, as substitutes for the usual statistical standards of reliability, two criteria will be employed, one of which is consistency of performance. Thus, if on a set of 15 items the Cui. sample should have one more correct answer than the Tra. sample on every example, then the former sample could be said to have consistently surpassed the latter.

But the consistent difference of one correct answer for 15 items represents only 6.7 percent of the possible score, and so, too little of a difference to affect educational practice. Hence, to the criterion of consistency a second must be added: the differences between mean performances should be of reasonable size. In Part III the quantitative standard set--arbitrarily again--is 10 percent of the possible score. There is just no sense in exaggerating the educational importance of very small differences between the scores of competing groups of school children.

CHAPTER V

RELATIVE ACHIEVEMENT ON SKILLS TAUGHT IN BOTH PROGRAMS; THE COM. TEST

The Com. Test (a copy of which will be found in the appendix) was devised to measure learning with respect to computational skills taught in the Cui. and in the Tra. programs alike. It is therefore an achievement test for both groups of subjects and should be equally valid for each sample since every item had been approved by panels of experts on the two programs. Scores on parts of the test representing sets of relatively homogeneous items are reported in Tables 11 to 17 in terms of Ac. (number of correct answers) and NAs (non-attempts). To assist in interpreting scores for the sets as wholes, data will be drawn from the item-by-item analyses for each set.

Set 1. The Simple Number Combinations.[1]

Table 11

a. The set as a whole. Items 1 - 20. Columns (4) and (7).

---

[1] In American mathematical programs the simple number combinations are ordinarily not viewed as computational items. In Scottish programs, however, they are so regarded. And it was of course the Scottish conception that was adopted for this study.

Table 11

The Com. Test. Scores of Samples on Simple Number Combinations

	Cui. Sample						Tra. Sample					
	Items 1 - 12		Items 13 - 20		Items 1 - 20		Items 1 - 12		Items 13 - 20		Items 1 - 20	
Scores	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA
(1)	(2)		(3)		(4)		(5)		(6)		(7)	
20					20	0					12	0
19					21	0					21	0
18					16	0					18	0
17					9	0					15	1
16					10	0					10	0
15					9	0					6	0
14					9	0					10	1
13					6	0					5	1
12	46	0			8	1	46	1			5	0
11	20	0			3	0	18	0			2	0
10	9	0			2	1	14	1			4	0
9	19	0			3	1	18	0			1	2
8	8	0	33	1	1	2	8	0	20	2	3	1
7	9	1	26	0	1	0	3	0	30	2	1	1
6	2	1	23	1	1	3	2	3	25	0	1	2
5	5	1	17	0	1	1	4	1	15	1	2	1
4	2	4	13	2	0	5	2	2	8	2	0	4
3	0	4	4	2	0	2	0	2	8	3	1	3
2	0	5	1	9	0	6	1	3	5	4	1	4
1	0	10	1	11	0	15	1	4	2	22	0	18
0	0	94	2	94	0	83	3	103	7	84	2	81
Means	10.11	0.55	6.13	0.48	16.18	1.03	9.83	0.58	5.53	0.68	15.47	1.21
S.D.'s	1.84	1.31	1.77	1.20	3.51	2.24	2.79	1.84	2.22	1.57	4.43	2.86

None of the differences noted below is significant at either the 0.01 level or the 0.05 level: Items 1 - 12, Ac. and NA; Items 1 - 20, Ac. and NA.

In all, 20 number combinations were included in the Com. Test, combinations in which the largest sum is 17 and the largest product is 35. Two forms of number sentence were presented. First were 12 combinations in which answers were to be inserted in the more traditional manner, at the end of number sentences; e.g., 8 + 5 = ____, 17 - 9 = ____, and 2 x 9 = ____. Next came a group of eight exercises in which the missing numerals had to be entered in the first or an intermediate place in number sentences; e.g., ____ + 7 = 10 and 2 x ____ = 12.

According to the data in column (4) of Table 11, the Cui. mean of correct answers for the 20 items is 16.18 with an S.D. of 3.51. The corresponding measures for the Tra. sample are 15.47 and 4.43--column (7). The Cui. mean for NAs is 1.03 with an S.D. of 2.24, to be compared with the Tra. mean of 1.21 and S.D. of 2.86. The difference between the Ac. means is but 0.71, and between the NA means is only 0.18. Since all subjects in both samples completed Set 1, the frequencies of correct answers, wrong answers, and NAs are directly comparable.

Next, consideration may be given to the results of item-by-item analysis, results which, neither here for Set 1 nor for other sets, will be presented in tabular form. To have included tables would have been to increase materially the length of the manuscript, and needlessly so, for citation to significant findings in the analyses will serve all essential purposes.

The Tra. subjects wrote the larger number of correct answers for four of the 20 items in the set, but never by more than 4. The Cui. subjects excelled in accuracy on fifteen items, on five by 8 or more; on seven by 5 or less. Details will be considered below.

The Cui. and the Tra. samples made the same number of errors in attempted answers on two examples, and the Tra. sample had fewer incorrect answers on four others, but by differences of only 4 or less. The Cui. sample was freer of errors on fourteen of the twenty items, in two by 10 or more.

A study of negative responses (incorrect answers plus NAs) revealed that, on the average, for each error made by the Cui. subjects, 0.34 of an example was omitted. For the Tra. subjects the ratio is 1.0 to 0.29. Stated differently, the Cui. sample is charged with approximately three errors for each NA; the Tra. sample, with slightly fewer than three errors for each NA. Similar pairs of ratios were calculated for all other sets in the Com. Test on the chance that the relationship noted here, for the Tra. subjects to have the smaller ratio, might prove to be characteristic. These pairs of ratios will be assembled and discussed in the concluding section of this chapter.

b. Group 1. Items 1 - 12. Columns (2) and (5).

In these twelve items the missing numeral is to be inserted at the end of each number sentence. The Ac. means of the Cui. and the Tra. samples are 10.11 and 9.83, both with large S.D.'s; and the NA means are 0.55 and 0.58, respectively, again with relatively large S.D.'s.

To turn now to the data obtained by item analysis: the Cui. subjects were the more accurate in dealing with seven of the number combinations. Their largest advantage is 5 on three items; their average advantage on the seven is 4.0. On the four combinations in which they excelled in frequencies of correct answers, the Tra. margins are 4 or less.

The two samples agreed fairly closely on the relative difficulty of the twelve combinations. The same two were placed in the easiest third (four items); the same three in the middle third, and the same three in the most difficult third. The Cui. subjects had the smaller number of incorrect answers on seven combinations. On these seven their average difference in frequencies of errors compared with the Tra. sample's is 4.0, and in no instance amounts to more than 5.

It could be true, of course, that actual unlikenesses existed among the number combinations in the different operations. The twelve combinations comprised three in each operation, A (addition), S (subtraction), M (multiplication) and D (division). The totals of correct answers in each operation are:

	Cui.	Tra.
A	345	332
S	293	286
M	292	290
D	272	255

The largest difference between pairs of totals is 17 in D in favor of the Cui. sample; but this difference, to say nothing of the others, is untrustworthy and insignificant. The Cui. total, 272, represents 76 percent of their opportunities (272 divided by 3 x 120); and the Tra. total of 255, 71 percent of their opportunities. Any argument for superiority of the Cui. subjects must be based upon the consistency of their greater success in all four operations, but it must take into account the fact that the differences in totals are too small to signify much.

c. Group 2. Items 13 - 20. Columns (3) and (6).

The examples in Group 2, with missing numerals in the first or in an intermediate position in the number sentences, proved to be more difficult than the items in Group 1, as probably would be expected. In percentage the Cui. Ac. mean dropped from 84.3 in Group 1 to 76.6, and the Tra. Ac. mean, from 81.9 to 69.1. The Cui. Ac. mean for Group 2 is larger by 0.60, and the Cui. mean for NAs is smaller by 0.20 than the corresponding means of the Tra. subjects. These differences are of course unreliable.

Item analysis for Group 2 revealed that the Cui. totals for correct answers exceeded the Tra. totals in seven of the eight items, by 8 or more in five instances. Item 18, _____ - 2 = 7, and item 19, _____ + 2 = 7, were especially troublesome to the Tra. subjects, who had only 52 correct answers for the first and only 38 for the second.

The same four items were in the easier half and, of course, the other four in the more difficult half for both samples, though the ranking within halves varied slightly for the two samples.

The Cui. subjects made the fewer errors in finding answers for seven of the eight examples, on only two by 10 or more. Their average advantage is 8.4, as large as it is because of the extreme difficulty the Tra. subjects encountered with items 18 and 19, as noted above.

The Tra. sample consistently made negative responses more commonly than did the Cui. sample, by margins of 3 to 19. On this account, in addition to the results of other comparisons, the Cui. subjects seem to have surpassed the Tra. subjects, if but little, on Group 2.

It is questionable, however, that the Cui. subjects really knew the eight combinations in the eight examples any better than did the Tra. subjects. It was known at the time when the Com. Test was constructed that the latter subjects in general had had little or no experience with examples in the form of 2 x ____ = 6 and ____ - 7 = 5. Accordingly it was agreed that, before starting on the Com. Test, they should be given some teaching in connection with completing number sentences so expressed. In all likelihood, the amount of practice they had, perhaps ten minutes in all, was insufficient for them to acquire anything like the ease of dealing with these forms that was possessed by the Cui. subjects. For the Tra. children items 13 - 20 may not have measured knowledge of number combinations as such, but this knowledge plus the ability to unravel number relationships of a comparatively strange sort. If this hypothesis is sound, then the Tra. panel apparently accepted for the Com. Test items that weakened the case for their program.[2]

On the whole, the two samples appear to have been equal in ability to supply answers for the number combinations in the Com. Test. This judgment was not anticipated.

---

[2] Yet, this panel did include in the Tra. Test to be considered in a later chapter four items (items 49 - 52) in which the missing numerals in the number sentences all occur in the position just preceding the answer. As will be seen, the Tra. subjects did not do well on these items, a fact which may or may not be explicable as the result of their unfamiliarity with the form of statement. If so, then why were they put in the Tra. Test?

In the Tra. program the number combinations are taught directly, explicitly, and systematically. In the Cui. program they are supposedly learned through activities (a) in discovering number relationships by manipulating the Cui. rods and (b) in using these discoveries in dealing with practical quantitative situations, (c) with little or no formal drill on the combinations as such. The Cui. programs seem to have been successful, at least to a point. It must be remembered, however, that the number combinations selected for the Com. Test had to be approved by the Cui. panel, and the choices made are those emphasized in the Cui. program, so that other combinations taught in the Tra. program were excluded. Hence, in the judgment stated above, to the effect that the two samples were on a par with respect to proficiency with the number combinations, it was necessary to insert the limiting phrase "number combinations in the Com. Test." Left unanswered is the question whether the two programs were equally effective in teaching the whole gamut of number combinations.

Set 2. Supplying Numerals in Number Sentences; A and S. Table 12.

Three groups of items, each successive group more difficult than the preceding, make up a total of fifteen items in the set. In Group 1, the largest sum is 24; in

Table 12

The Com. Test. Scores of Samples in Completing Number Sentences, Addition and Subtraction

Scores	Cui. Sample								Tra. Sample							
	Items 21 - 25		Items 29 - 32		Items 56 - 61		Total Preceding		Items 21 - 25		Items 29 - 32		Items 56 - 61		Total Preceding	
(1)	Ac. (2)	NA	Ac. (3)	NA	Ac. (4)	NA	Ac. (5)	NA	Ac. (6)	NA	Ac. (7)	NA	Ac. (8)	NA	Ac. (9)	NA
15							0	1							0	2
14							11	0							3	0
13							8	4							4	1
12							3	5							7	6
11							9	4							10	2
10							7	4							9	4
9							6	7							12	7
8							13	3							9	6
7							14	8							6	14
6					5	34	3	13					0	35	14	10
5	46	3			14	9	11	6	38	6			8	11	8	6
4	30	3	21	16	9	6	10	4	35	2	13	14	17	8	11	8
3	22	2	31	7	9	3	7	5	13	2	27	4	12	4	5	8
2	7	8	20	14	11	11	5	8	11	8	26	17	12	7	6	7
1	11	13	20	19	17	11	11	17	12	23	22	29	18	16	10	13
0	4	91	28	64	55	46	2	32	11	79	32	56	53	39	6	26
Means	3.68	0.52	1.98	1.10	1.68	2.63	7.28	4.33	3.39	0.69	1.73	1.09	1.55	2.83	6.71	4.66
S.D.'s	--	--	--	--	--	--	4.10	4.24	--	--	--	--	--	--	3.87	4.01

None of the differences is significant between the Cui. and the Tra. means for Ac. and for NA (columns 5 and 9) at either the 0.01 or the 0.05 level.

Group 2, 37, and in Group 3, 64. Illustrative examples are $9 + 3 + \_\_\_\_ = 16$, $37 = 25 + \_\_\_\_$, and $\_\_\_\_ - 13 = 13$.

The numbers of items in the groups are five, four, and six,--clearly too few to justify generalizations for any group. As a matter of fact, generalizations for the set as a whole are made hazardous by the fact that many subjects in both samples did not have time to try the computations called for in the last six examples.

a. The set as a whole. Items 21 - 25, 29 - 32, and 56 - 61. Columns (5) and (9).

The Cui. and the Tra. means for accuracy (7.28 and 6.71, respectively) differ little, as is true for the NA means (4.33 and 4.66). That the Tra. subjects did as well as they did in comparison with their rivals was somewhat surprising in view of their lack of experience with examples written in the form of horizontal sentences, as noted in Section 1 above.

It is hoped that by now the reader has become accustomed to the pattern of organization for the discussion of results on each set or group of items. The data in the basic table (here, Table 12) having been considered, a transition is made to the findings of item analysis.

The Cui. totals of correct answers for Set 2 surpassed the Tra. totals in twelve of the fifteen examples. The largest difference is 21, in example 24, 18 = 7 + ____ + 6. In terms of correct answers the five examples most difficult for the Tra. subjects were also most difficult for the Cui. subjects, and in precisely the same order of ranks. And another five examples proved to be the easiest five for both samples, with but two of these items differing in rank order.

The Cui. sample made fewer errors than the Tra. sample in ten of the fifteen examples, in only two by 10 or more incorrect answers. The Tra. subjects had fewer errors in three examples by even smaller margins. For the other two examples the numbers of erroneous answers is the same.

b. Group 1. Items 21 - 25. Columns (2) and (6).

On the average the Cui. subjects secured the larger number of correct answers for 0.29 more examples and attempted 0.17 more examples.

All subjects finished Group 1. The Cui. subjects had more correct answers in four examples; the Tra. subjects, in only one. A single difference is larger than 7, that (in favor of the Cui. sample) being for item 24, already referred to. The five examples have substantially the same ranking for difficulty for both samples as judged by numbers of correct answers.

The Cui. sample made the fewer errors in all five examples, in none by more than 6.

c. Group 2. Items 29 - 32. Columns (3) and (7).

Among the four items in this group is one that is odd, in that it calls for knowledge that 3 doz. = 36. The Ac. means of the samples differ by only 0.25 in favor of the Cui. subjects. The two NA means are practically identical.

The greater frequencies of correct answers were made by the Cui. subjects in all of the four items. The amount of the differences in this respect are never larger than 7. The four examples rank in the same order of difficulty for both samples.

In their attempts to secure answers for three of the examples the Cui. subjects made the fewer errors, in two of them by 10 and 12.

d. Group 3. Items 56 - 61. Columns (4) and (8).

On the average the Cui. subjects outscored the Tra. subjectsby 0.13 in Ac. and omitted 0.20 fewer examples. For four items the frequencies of correct answers for the Cui. subjects exceeded those of the Tra. subjects by 9 or fewer. The difficulty ranking of the examples, in terms of numbers of correct answers, is substantially the same for both samples. As judged by totals of negative responses, items 60 and 61 (____ - 13 = 13, and ____ + 15 = 51) proved to be the most troublesome for both samples, but for one reason because of the large increases in NAs, owing to shortage of time.

In conclusion, there seems to be no reliable and so, no significant difference in the ability of the two samples to deal with the items in Set 2, either in Ac.[3] or in NAs. True, on the set as a whole, the Cui. subjects (a) had more correct answers than the Tra. subjects for twelve of the fifteen examples, by an average amount of 7.8; (b) had fewer incorrect answers on nine of the items, by margins of 6 or more on five, and (c) is charged with fewer NAs on ten examples, by 7 or more on only five. Yet, the Cui. advantages, while fairly consistent, for the most part are small.

In items like those in Set 2 there is some interest in knowing whether the relative difficulty of examples was affected by the place in the number sentence where the missing numeral is to be inserted. As judged by totals of negative responses, the order of difficulty for both samples was the same, easiest to most difficult: intermediate, terminal, and initial, though the difference in totals for the easier two places is slight. Moreover, the number of items in each group is small: initial, 4; intermediate, 6; terminal, 5. Yet again, the special difficulty of the group in which the missing numeral occurs first in the number sentence may be an artifact. Two of the four items in this group came very late in the test and were therefore subject to an excessive number of NAs.

---

[3] A different method of comparing the accuracy of the two samples (that is, different from those reported above) was tried out for Set 2. The subjects who did not start Group 2 were eliminated in consideration of the items in that group as well as in Group 3; and the subjects who did not reach the first item in Group 3 were eliminated in consideration of the Group 3 items. There were left, then, only the subjects who attempted all items in the group or groups, as well as those who omitted apparently difficult items as a matter of choice. Percentages of correct answers were then computed for both reduced samples on each item. Results of the ensuing comparisons are not here reported, for they changed the general picture not at all.

Set 3. Horizontal and/or Vertical Addition: Three Addends. Items 33, 34, 62 - 64. Table 13.

This set comprises only five items. Each was printed in both the horizontal and the vertical forms of presentation; thus,

$$19 + 26 + 34 = \underline{\qquad} \;\; \underline{\text{or}} \;\; \begin{array}{r} 19 \\ 26 \\ \underline{+\ 34} \\ \\ \underline{\qquad} \end{array} \quad . \text{ To both samples}$$

it was explained prior to the testing that each child could choose the form he preferred; and, besides, the phrase "Add either way" appeared on the test blank twice. The five items were offered in two separated groups, 33 and 34; 62 - 64. There being so few examples in all, they will be considered as a single set.

Table 13

The Com. Test. Scores of Samples in Vertical and/or Horizontal Addition

Scores	Items 33, 34, 62 - 64			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
5	27	11	39	8
4	19	4	22	3
3	16	24	8	24
2	20	2	18	6
1	13	5	6	2
0	25	74	27	77
Means*	2.60	1.27	2.91	1.15

*S.D.'s omitted as meaningless.

The Cui. sample made 111 percent of the Tra. number of negative responses. Four of the Cui. subjects and five of the Tra. subjects stopped work before coming to the first item in the set, and an additional 29 of the former and an additional 26 of the latter, before coming to the last three items, which appeared fairly late in the test. Thus, 107 of the total of 153 Cui. NAs were caused by shortage of time, and 120 of the total of 138 Tra. NAs were similarly caused. Hence, the Cui. subjects skipped eight more examples[4] than did the Tra. subjects, examples thought to be too difficult and hence subject to unwanted errors.

---

[4] Total of NAs minus total of NAs resulting from lack of time.

All in all, while no measure of reliability can be calculated, the results of the comparisons above give the impression that the Tra. sample excelled in horizontal/vertical addition. Nevertheless, the small margin of advantage enjoyed by this sample is threatened by an extraneous factor; namely, the confusion of many Cui. subjects occasioned by the presentation of the same examples in the two forms of expression. The Tra. subjects consistently added by columns. Sixteen Cui. subjects did so, but only one was successful with all five examples. Work on the test papers of 32 Cui. subjects reveals uncertainty,--whether to add horizontally or to add by columns. Some started by using the horizontal form and changed to use of columns. Others used the horizontal form in an example and then copied the answer, correct or incorrect, under the corresponding vertical form. Still others varied throughout the set, now adding horizontally and now vertically.

The inability of about a fourth of the Cui. subjects to decide what was to be done in Set 3 was most probably caused by inadequate explanations of the direction, "Add either way," which were given before the actual testing began. In any case inability to decide what to do has little relationship to the purpose of Set 3, which was to measure skill in horizontal/vertical addition. The effect, however, was almost certainly to lower the Ac. scores of the 32 Cui. subjects referred to in the foregoing paragraph, and probably to lower them enough to wipe out the apparent superiority of the Tra. sample. This possibility, in addition to the fact that almost a fourth of the subjects in each sample lacked time to attempt the last three items in the five-item set, leads one to conclude that on this set neither sample overmatched the other.

Set 4. Fractions. Items 26 - 28, 35 - 37, 55.
Table 14

The first three items are of the type, find $\frac{1}{2}$ of 20 and divide by 5; the last four, of the type, What is $\frac{1}{2}$ of 86?

For the seven examples combined the two means differ by only 0.07 in Ac. (in favor of the Cui. subjects) and by only 0.05 for NAs (in favor of the Tra. subjects). (Table 14.)

So much for the data in the basic table; now to turn to the results of item analysis: the Cui. subjects have an average of 4.5 more correct answers in four examples; the Tra. subjects, an average of 7.0 more in three. Both samples found item 35 (What is the half of 16?) the easiest in the set and agreed on the same two other examples as next in difficulty, but with relative ranks reversed. Item 28 was the most troublesome for both samples,--Multiply $\frac{1}{2}$ of 16 by 3.

Items 26 - 28 are more complicated than are items 35 - 37 and 55. In each of the former examples an answer has to be known or found for such a fractional expression as $\frac{1}{2}$ of 16, and then the answer has to be used in making a computation. Finding the answers for the second group of items involves, or may involve, no more than knowledge of fractional parts that could be memorized. Being more complicated, the first three examples discussed might be expected to be more difficult than were items 35 - 37 and 55; and they were, for both samples had identical average percentages of 46 for the first three and of 75 (Cui.) and 54 (Tra.) for the last four.

The distributions of scores for the two samples in Table 14 are very similar. Forty-five Cui. and 44 Tra. subjects made Ac. scores of 5 to 7; 44 Cui. and 40 Tra. subjects, scores of 0 to 2.

The Cui. subjects made the smaller numbers of errors in four examples and the larger in three. One difference amounts to 16 (Cui. for item 27), and none of the other six, to more than 6. In each sample three subjects stopped work before reaching item 35; 19 more Cui. and 15 more Tra. subjects did not have time to try item 55.

Table 14

The Com. Test. Scores of Samples on Fractions.

Scores	Items 26 - 28, 35 - 37, 55			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
7	13	7	13	9
6	23	4	16	5
5	9	6	15	2
4	19	9	20	9
3	12	11	16	7
2	20	8	11	9
1	13	24	17	28
0	11	51	12	51
Means	3.66	1.77	3.58	1.72
S.D.'s	2.23	2.17	2.21	2.22

To sum up, the two samples appear to have done equally well in their work with fractions, save for the poorer record of the Tra. sample on the last four items of Set 4. Yet, there were large differences in the demands made upon the two samples by reason of dissimilarities in the programs they had studied. Because of greater experience in halving and doubling numbers, the Cui. subjects had an advantage that gave them special assistance in all seven examples. They, or many of them, may well have memorized the answers for

$\frac{1}{2}$ of 12, of 20, of 16 (which occurred twice), and of 86, as well as 1/5 of 30. By contrast, in most of these instances, at least for most Tra. subjects, the answers had to be found through "short division." (See Mr. Allan's statement, page 22 of Chapter I.) The Tra. sample, therefore, did surprisingly well to equal the Cui. sample.

Set 5. Structure and Divisibility of Numbers.
Table 15.

a. The set as a whole. Items 46 - 48, 70 - 72.
Columns (4) and (7).

The first three items relate to number structure, and typical of this group is the example, Write these numbers in figures: Six hundred and five. The last three items are concerned with the divisibility of numbers; and typical of the last three is the example, Write three numbers that divide into 20 exactly.

The means of the two samples for the six items together are identical (2.72) for Ac., and the Cui. mean for NAs is only 0.21 the smaller. But these statistics, as well as others to follow, cannot be taken too seriously because of the large numbers of work stoppages,--in the Cui. sample, eight before reaching item 46 and 50 more (58 in all, or nearly half) before reaching item 70; in the Tra. sample, 11 before reaching item 46 and 53 more (64 in all, or more than half) before reaching item 70.

Table 15

The Com. Test. Scores of Samples on Number Structure and Number Divisibility

Scores	Cui. sample						Tra. Sample					
	Items 46 - 48		Items 70 - 72		Total Preceding		Items 46 - 48		Items 70 - 72		Total Preceding	
	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA
(1)	(2)		(3)		(4)		(5)		(6)		(7)	
6					14	10					8	16
5					8	3					11	0
4					6	5					1	2
3	75	14	15	58	50	45	79	16	9	66	63	51
2	10	2	9	12	9	9	14	0	10	3	13	6
1	11	3	10	8	13	10	4	2	4	18	4	14
0	24	101	86	42	20	38	23	102	97	33	20	31
Means	2.13	0.41	0.66	1.72	2.72	2.15	2.34	0.42	0.43	1.85	2.72	2.36

For whatever they are worth, the data show that each sample had the greater number of correct answers for three examples, by an average of 5.3 for the Tra. subjects and of 8.3 for the Cui. sample. Starting with the easiest example, five examples have the same ranking for difficulty in both samples. In each of four examples the Tra. subjects made the fewer errors by an average of 5.2; in each of two, the Cui. subjects by an average of 2.

b. Group 1. Number structure. Items 46 - 48.
Columns (2) and (5).

All except eight Cui. and 11 Tra. subjects completed the examples in Group 1; but since only three items were to be dealt with, the significance of the findings on "number structure" is slight and uncertain. The Tra. mean for accuracy is a bit the larger of the two; and the difference in average number of errors per example is 5.3, in favor of the Tra sample. In all three examples the accuracy of the Tra. sample was greater than that of the Cui. sample, but by as many as 8 correct answers on only one item. The two means for NAs are practically identical, as is the order of difficulty of the examples for the two samples as judged by totals of correct answers.

c. Group 2. Number divisibility. Items 70 - 72. Columns (3) and (7).

Comparisons of the pairs of means for numbers of correct answers per example and for numbers of NAs per example favor the Cui. subjects (differences of 0.23 and 0.64, respectively). The numbers both of errors and of NAs per item are much larger for Group 2 than are the corresponding numbers for Group 1. On all three items the Cui. subjects had the greater number of correct answers (an average of 8.7), and on two of the items the smaller numbers of errors. In view of the excessive amount of work stoppage because of time limits prior to item 70, the data on "number divisibility" are very incomplete and are therefore not reported in the usual manner.

To add up, the slight evidence available might point to superiority of the Tra. subjects on "number structure," and of the Cui. subjects on "number divisibility," but one can only surmise. If the tentative inferences were warranted, it would be in line with what is known about the two programs, number structure being emphasized more in the Tra. program, and number divisibility in the Cui. program.

Set 6. Two-operation Examples. Items 49 - 54.
Table 16

The first four of the six examples are of the type, 34 - 25 + 8 = ____, and the last two items, of the type, Multiply 10 by 5 and take away 24 ____.

Again the records of the two samples indicate parity in the abilities tested. The Tra. subjects' mean for Ac. is the greater by 0.10, and their mean for NAs is the smaller by 0.25. On three items the Tra. sample averaged 8.3 more correct answers; on two, the Cui. sample by an average of 1.5. The Tra. subjects made 10 more errors on the last item, 23 + 41 - 32 = ____, than did the Cui. subjects, who had fewer errors in three other examples by differences of 4 or fewer. The relative difficulty of the six items was similar for the two samples. The same three examples appeared in the easier half, and another three in the more difficult half, with but minor differences in ranks within the halves.

Fourteen Cui. subjects and 21 Tra. subjects did not reach the first item in Set 6. Despite the difference of seven subjects no correction has been made in the data reported above.

Table 16

The Com. Test. Scores of Samples on Two-operation Examples

Scores	Items 49 - 54			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
6	5	15	6	15
5	8	12	11	3
4	14	6	11	6
3	13	3	16	10
2	17	14	14	11
1	31	8	31	8
0	32	62	31	67
Means	1.92	1.83	2.02	1.58

Finally one may conjecture why the Cui. subjects did not considerably outscore the Tra. subjects on Set 6, more especially on the first four items. Certainly they (the Cui. subjects) had had a great deal of experience with two-operation examples. (See the Cui. Test.) But for them, such examples were commonly, if not generally, expressed with brackets or parentheses. Thus, item 51 of the Com. Test, 66 - 30 - 8 = ____, would very probably have been presented in some such fashion as (66 - 30) - 8. Perhaps the absence of the familiar cues of parentheses (or brackets) served to confuse them, leaving them uncertain concerning the point at which to start computation. On the other hand, the Cui. panel raised no objection to the inclusion of the items in question in the Com. Test; nor did it request that an explanation of the form of the items be given the Cui. subjects prior to the administration of the test.

Set 7. Time measures; Scottish money.
Items 38 - 45, 65 - 69. Table 17.

The Ac. means of both samples are low, representing percentages of less than 50; and the NA means are large, representing more than 30 percent of the possibilities. The Cui. Ac. mean exceeds the Tra. mean by a mere 0.15, and the Cui. subjects on the average had 0.39 fewer NAs. Of the most difficult five examples for the Cui. sample, four were in the most difficult five for the Tra. sample, and another five examples were the easiest five for both samples.

The Cui. totals of correct answers are the larger for eight examples, in only three by 8 or more (10, the largest); the Tra. totals, the larger in five, in two by 8 and 14. The placement of items 65 - 69 near the end of the test was largely responsible for a total of 472 NAs for the whole set in the case of the Cui. sample and a total of 528 NAs in the case of the Tra. sample. If comparisons for accuracy are limited to the earlier items in the set, 38 - 45, the Cui. average number of correct answers per example is 66.6, and the Tra. average is 65.4.

Table 17

The Com. Test. Scores of Samples on Time Measures and Scottish Money

	Items 38 - 45, 65 - 69			
Scores	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
13	2	8	1	11
12	2	3	6	3
11	10	2	8	1
10	6	1	5	1
9	10	3	7	2
8	9	4	14	7
7	11	5	11	5
6	6	9	10	8
5	15	17	8	19
4	16	9	10	6
3	7	8	8	8
2	8	2	6	3
1	6	10	8	12
0	12	39	18	34
Means	5.64	3.93	5.49	4.32
S.D.'s	3.54	4.04	3.79	4.19

There is no significant difference, either at the 0.01 or at the 0.05 level, between the Cui. and the Tra. means for Ac. or for NA's.

Since time measures were used in only two items, the relative achievement of the two samples is of but incidental interest. Item 68 is: 1 hour - $\frac{1}{2}$ hour = _____ minutes. The frequencies of correct answers are 29 for the 60 Cui. subjects who attempted the item and 32 for the 55 Tra. subjects who tried the computation. Item 69 is: 50 minutes and 20 minutes will make _____ hours and _____ minutes.

Thirty-three Cui. subjects and 37 Tra. subjects were successful out of the 59 Cui. and the 55 Tra. subjects who worked on the example.

In sum, because of the small differences between Ac. and NA means and because of the excessively large number of NAs beginning with item 65, it is impossible to say that either sample surpassed the other on this set of 13 items. Had Set 7 contained triple the number of items and had all subjects finished the longer test, differences in ability to deal with Scottish money and with measures taught to both samples might have been revealed. However, the only evidence available is the performance of the two samples on a short test, in which they seem to have done equally well.

Summary and Interpretation

This chapter reports the results of comparing the records of the Cui. and the Tra. samples, each comprising 120 subjects, on the Com. Test which, because of the method of its development, was an achievement test for both the Cui. and the Tra. subjects. But interest lies, not in the relative success of the two samples on the test as a whole,[5] but rather in their success on types of computational skills.

---

[5] As shown in Table 10, the Ac. mean of the Cui. sample was 1.45 greater than that of the Tra. sample (40.28 compared with 38.83). The difference is unreliable.

The 72 items in the test were distributed into seven sets of comparatively homogeneous items; e.g., computations with fractions. These sets contain from five to 20 items; and three of the sets were broken down into sub-sets or groups in order to recognize minor differences in homogeneity. The individual test papers were examined item by item, as well as by sets and groups, to identify possible dissimilarities in which one or the other of the two instructional programs might prove to have been more effective in developing computational proficiency.

The relevant data are assembled in Table 18; but before the data are scrutinized it is important to remember that even the most complete set consists of 20 items; the next most complete, of 15 items, and the remaining five, of 13 items or fewer. And the largest group in any set comprises only 12 items. When samples of test items number 10 or fewer items, statistical measures of reliability as they are customarily calculated are highly suspect.

Hence, in the interpretation of differences between the scores of the two samples on sets and groups of items, two criteria of reliability will be employed. One criterion will be consistency. This term means an uninterrupted succession of advantages of one sample over the other in a series of at least six comparable items. But the margins of advantage must also be of reasonable size, and this is the second criterion. Reasonable size will be defined as margins equal to at least 10 percent. To illustrate: if the Tra. sample should have but one more correct answer than the Cui. sample on each of the items in a set of 13, the criterion of consistency would be satisfied but not the criterion of reasonable size.

As might be expected, the achievement of schools in each sample, Cui. and Tra., are very uneven. In the Cui. set of schools three made outstanding records and two, very poor records. In the Tra. set of schools were two with very inferior records and two with superior records. The marked differences in achievement cannot be explained as the result of large differences in intelligence ratings. Rather, the differences in performance on the Com. Test reflect unlikenesses within each set of schools in the degree to which they adhered to the particular program they were purported to follow. In a word, it is a myth to refer to the Cui. program and the Tra. program. This statement merely underscores the same sort of statement made by Miss Law and by Mr. Allan in Chapter I, where both stress the lack of uniformity of schools supposedly teaching the same program.

Review of the results by sets

Set 1. Number combinations; 20 items. The two samples were equally accurate on the eight combinations in which answers were to be written in the last position in number sentences. On the other twelve items, in which missing numerals were to be inserted in the first position or at the middle of number sentences, the Tra. subjects were regularly inferior; but their difficulty seemed to lie less in ignorance of number combinations than in the necessity of dealing with them in a form unfamiliar to them.

Set 2. Completing number sentences; 15 items. Illustrations of three types of item are: 9 + 3 + ____ = 16, ____ = 14 + 21, and 64 - ____ = 11. All items call for addition and/or subtraction, and nine involve four numerals. The record of the Cui. sample is slightly the better of the two, but the margins of advantage are small. Again the Tra. subjects were at something of a disadvantage because of relatively less prior experience with some of the forms of the examples.

Set 3. Vertical and/or horizontal addition; five items. Each of the five examples was printed both in the horizontal and in the vertical forms, preceded by the instruction, "Add either way," and with the word "or" inserted between the pairs of algorisms. Each example has five addends, and three examples have two-place numerals; sums are all less than 99. The Tra. subjects outscored the Cui. subjects in accuracy on all five examples, but by amounts of small size. About one fourth of the Cui. subjects were handicapped by failure to understand directions as shown by their work on their test papers.

Set 4. Fractions; seven items. Three of the examples are of the type, Add 20 to $\frac{1}{2}$ of 12; four of the type, What is the half of 82? Neither group surpassed the other in accuracy on the set as a whole, though there is a strong probability that the Tra. subjects had had less instruction on the skills involved.

Set 5. Number structure and divisibility; six items. On the three items of the type, Write this number in figures: Three Hundred and forty-five,--on these items the Tra. sample did slightly better than the Cui. sample. The other three items in the set, illustrated by, Write three numbers that divide exactly into 24, came last in the test. NAs were so common that no comparison of relative success is warranted.

Set 6. Two-operation examples; six items. Samples of a group of four items are 12 + 26 - 30 = _____ and 34 - 25 + 8 = _____. One of the two other items is, Find 2 times 8 and add 10. The slight superiority of the Tra. sample in frequencies of correct answers was in part produced by a factor extraneous to the purpose of the testing. Work on test papers for the first four examples showed that many Cui.

subjects were confused by the form of statement. They were accustomed to seeing such examples written with brackets, and in their absence they did not seem to know where to start the computations.

Set 7. Time measures; Scottish money; 13 items. Only two of the items involve time measures. The Cui. subjects have the higher average accuracy, but by small margins. The fact that on each of the last five items, which come near the end of the test, there were more than 60 omissions for each sample makes it impossible to say that either sample was certainly superior to the other.

Review of the results in terms of accuracy means. Table 18.

The statistics in columns (2) and (3) of Table 18 were obtained by comparing the Cui. and the Tra. means in Tables 11 to 17. For example, according to Table 11 the Ac. means (average number of correct answers) for Set 1 as a whole are: Cui., 16.18; Tra., 15.47. The difference between the two is 0.71 in favor of the Cui. sample and is so entered in the first row of data in Table 9a. In similar fashion the NA means (average number of omissions) are: Cui., 1.03; Tra., 1.20. The difference, 0.18, in favor of the Cui. sample (because of fewer omissions), appears as the first entry in column (3).

Table 18

Summary of Data for Sets and Groups of Computational Skills

Skills, with numbers of items in each (1)	Differences between Means,* with Superior Sample		Numbers of Individual items in which each sample had the Larger Number of Correct Answers (4)	
	Correct Answers (2)	Non-attempts (3)		
1. Number combinations				
Whole set (20)	Cui., 0.71	Cui., 0.18	Cui., 15	Tra., 4
Group 1 (12)	Cui., 0.18	Cui., 0.03	Cui., 7	Tra., 4
Group 2 (8)	Cui., 0.60	Cui., 0.20	Cui., 8	
2. Supplying missing numerals				
Whole set (15)	Cui., 0.57	Cui., 0.33	Cui., 12	Tra., 2
Group 1 (5)	Cui., 0.29	Cui., 0.17	Cui., 4	Tra., 1
Group 2 (4)	Cui., 0.25	Tra., 0.03	Cui., 4	
Group 3 (6)	Cui., 0.13	Cui., 0.20	Cui., 4	Tra., 1
3. Vertical and/or horizontal addition (5)	Tra., 0.31	Tra., 0.12	Tra., 5	
4. Fractions (7)	Cui., 0.08	Tra., 0.05	Cui., 4	Tra., 3
5. Number structure and divisibility (6)	0.00	Cui		
Whole set (6)	0.00	Cui., 0.21	Cui., 3	Tra., 3
Group 1 (3)	Tra., 0.21	Cui., 0.02	Tra., 3	
Group 2 (3)	Cui., 0.23	Cui., 0.13	Cui., 3	
6. Two-operation examples (6)	Tra., 0.10	Tra., 0.25	Cui., 2	Tra., 3
7. Measures; Scottish money (13)	Cui., 0.15	Cui., 0.39	Cui., 8	Tra., 5

*Derived by comparing means in Tables 11 to 17.

Examination of the data in column (2) of Table 18 shows that the Cui. sample surpassed the Tra. sample in average number of correct answers per total set in four instances, and the Tra. sample surpassed the Cui. sample in two total sets. The Cui. means are the greater by 0.71 in Set 1 and 0.57 in Set 2, but by 0.15 or less in Sets 4 and 7. The differences in Sets 3 and 6, in which the Tra. sample excelled, are 0.31 and 0.10. None of the differences is large enough to be educationally significant.

The entries in column (4) indicate little consistency for either sample in securing correct answers for individual items within sets. Thus, in Set 1 the Cui. subjects made the more successful attempts with 15 number combinations and the fewer with four. In Set 3 only did one sample (the Tra. subjects) uniformly surpass the other. But in this instance the differences in frequencies of correct answer item by item are small, and an explanation for the poorer record of the Cui. subjects has been offered just above.

In four sets the Cui. sample averaged the fewer NAs, and in three sets the Tra. sample enjoyed this distinction. (The entries in column (3) are read negatively, as it were. Cui., 0.18 in the first row means that the Cui. subjects on the average omitted 0.18 fewer items in Set 1.) This matter of NAs will be considered later.

On the whole, the Cui. subjects seem to have a slight edge over the Tra. subjects in achievement on the Com. Test, but an edge of insufficient size to justify the general adoption of either of the two programs to the exclusion of the other. In the case of sets in which the Cui. sample, or the Tra. sample, shows to advantage, more or less extraneous factors mentioned earlier in the chapter and to be cited again shortly, may in themselves account for such disparities as appear to exist.

In sum, neither of the foregoing analyses of records made on the seven types of computational skill included in the Com. Test reveal a single skill in which one of the samples clearly excelled. Hence, in all probability the two programs were equally effective in promoting computational competence. The conclusion is restricted of course to achievement on selected skills, and more particularly on these skills as they were used in the particular examples in the test, examples taught in both programs in the opinion of the Tra. and the Cui. panels. Moreover, the conclusion of equal effectiveness is subject to modification in terms of the results of other analyses to be reported.

Errors and Non-Attempts

Obviously on each example a child could have (a) recorded the correct answer, (b) written an incorrect answer, or (c) not attempted the computation at all. The data on

(a) have been reported in the foregoing section; but it is important to know about (b) and (c) also.

Errors. By sets and examples the data on errors are assembled in Table 19. In Set 1, for example, the Cui. subjects made fewer errors in 14 examples; the Tra. subjects in four. The Cui. sample made fewer errors by 8 or more in five examples; by 5 or less in nine; the Tra. sample, fewer errors by 4 or less in four examples.

In six of the seven sets each sample made both the greater and the smaller number of errors on one or more examples. In other words, save in Set 3 where the Tra. sample made the smaller number of errors on all examples, neither sample consistently was the more inaccurate in any set.

True, the Cui. subjects made the smaller number of errors in five sets, compared with two sets for the Tra. subjects; but a scrutiny of the data in the last column reveals that only in a few instances did the difference in errors amount to much. Even a difference of 10 errors on an example is equivalent to but 8.07 percent of the possibility,--120 (subjects) divided by 10 (errors).

The results of these comparisons do not seem to require any alteration of the conclusion that neither program was more effective than the other in producing computational proficiency in skills taught in both.

Table 19

Differences between Samples in Numbers of Errors, by Sets and Examples

Sets, with Numbers of Items	Numbers of Examples in which Samples made fewer Errors. Cui.	Tra.	Margins of differences in these examples
1. (20 items)	14	4	Cui.: fewer errors by 8 or more in 5 examples; by 5 or less in 9 Tra.: fewer errors by 4 or less in 4 examples
2. (15 items)	10	3	Cui.: fewer errors by 10 and 12 in 2 examples; by 6 or less in 8 Tra.: fewer errors by 6 or less in 3 examples
3. (5 items)	0	5	Tra.: fewer errors by 8 in 1 example; by 6 or less in 4 examples
4. (7 items)	4	3	Cui.: fewer errors by 16 in 1 example; by 5 or less in 4 examples Tra.: fewer errors by 6 or less in 3 examples
5. (6 items)	2	4	Cui.: fewer errors by 1 and 3 in 2 examples Tra.: fewer errors by 7 and 8 in 2 examples; by 4 or less in 2 examples
6. (6 items)	4	1	Cui.: fewer errors by 10 in 1 example; by 4 or less in 3 examples Tra.: fewer errors by 1 in 1 example
7. (13 items)	7	4	Cui.: fewer errors by 11 in 1 example; by 7 or less in 6 examples Tra.: fewer errors by 12 in 1 example; by 8 or less in 3 examples

Non-Attempts. For the sake of convenience and ease of reference, data are copied below from Tables 11 to 17. The measures represent averages of NAs per set.

Sets	Cui.	Tra.	Sets	Cui.	Tra.
1.	1.03	1.21	5.	2.15	2.36
2.	4.33	4.66	6.	1.83	1.58
3.	1.27	1.15	7.	4.04	4.19
4.	1.77	1.72			

The largest of the differences between the seven pairs of means is 0.33 (Set 2), and all others are 0.25 or less. The Cui. subjects have the smaller average for four sets; the Tra. subjects, for three. There is still no evidence to contradict the conclusion of equal effectiveness of the two programs in teaching the skills in the Com. Test.

Ratios of errors to NAs. In the discussion of the results for Set 1 it was noted that the Cui. subjects appeared to be slightly more prone than were the Tra. subjects to omit seemingly difficult examples rather than to risk errors. At this point the evidence bearing on this hypothesis will be examined; and the pertinent data are summarized in Table 20.

First, however, it is to be noted that NAs may be of two kinds: (a) omissions because of difficulty or supposed difficulty, and (b) omissions caused by lack of time. No distinction was made between the two kinds in the tabulation just above. In these paragraphs, however, as a second step omissions of type (b) were excluded, and the reduced totals of NAs are compared with the totals of errors for each set as a whole, in the form of ratios, errors to NAs.

In Set 1 all subjects in both samples attempted the 20 items, so that no adjustment needed to be made of the kind mentioned just above. Or, stated differently, the unadjusted (or "raw") ratios can be accepted as adjusted ratios. For Set 2 and all subsequent sets both unadjusted and adjusted ratios are reported. In Set 2, 60 NAs were eliminated for the four Cui. subjects who did not start the set of 15 items, and 156 more NAs were excluded for the 26 more Cui. subjects who did not have time to work on the last six items. This explanation illustrates the procedure employed to secure the numbers of NAs in other sets attributed to willful omission.

The ratios in Table 20 give little support to the hypothesis in question. In five of the six unadjusted ratios those for the Tra. sample are the smaller, but by slight differences only, save for Set 3. In the case of adjusted ratios, which probably are more significant in the present context, the Tra. ratios are the smaller in only four of the seven comparisons, and the differences are very small indeed. The initial hunch seems to have been invalidated, and it is

3

Table 20

Ratios of Numbers of Errors to Numbers of Non-Attempts, by Sets as Wholes

Sets	Unadjusted Ratios	Adjusted Ratios
1.	- - - - - - - - - -	Cui.: 1.0 : 0.34* Tra.: 1.0 : 0.29*
2.	Cui.: 1.0 : 1.17 Tra.: 1.0 : 1.21	Cui.: 1.0 : 0.77 Tra.: 1.0 : 0.89
3.	Cui.: 1.0 : 2.05 Tra.: 1.0 : 1.22	Cui.: 1.0 : 1.27 Tra.: 1.0 : 0.34
4.	Cui.: 1.0 : 1.22 Tra.: 1.0 : 1.00	Cui.: 1.0 : 0.87 Tra.: 1.0 : 0.74
5.	Cui.: 1.0 : 1.80 Tra.: 1.0 : 1.26	Cui.: 1.0 : 0.40 Tra.: 1.0 : 0.43
6.	Cui.: 1.0 : 0.80 Tra.: 1.0 : 0.65	Cui.: 1.0 : 0.50 Tra.: 1.0 : 0.42
7.	Cui.: 1.0 : 1.18 Tra.: 1.0 : 1.16	Cui.: 1.0 : 0.56 Tra.: 1.0 : 0.52

*Set 1. All finished the set; hence, no unadjusted ratio.

Set 2. Cui. sample: 4 subjects did not reach Group 2; 30, Group 3.
Tra. sample: 4 subjects did not reach Group 2; 26, Group 3.

Set 3. Cui. sample: 4 subjects did not reach the first example in the set; 33, the last 3 items.
Tra. sample: 5 subjects did not reach the first example in the set; 30, the last 3 items.

Set 4. Cui. sample: 3 subjects did not reach item 26; 6, item 35; 25, item 55.
Tra. sample: 3 subjects did not reach item 26; 6, item 35; 21, item 55.

Table 20 (Continued)

Set 5. Cui. sample: 8 subjects did not reach item 46; 58, item 70.
Tra. sample: 11 subjects did not reach item 46; 64, item 70.

Set 6. Cui. sample: 14 subjects did not reach item 49.
Tra. sample: 21 subjects did not reach item 49.

Set 7. Cui. sample: 7 subjects did not reach item 38; 39, item 65.
Tra. sample: 5 subjects did not reach item 38; 19, item 65.

---

to be discredited on logical as well as on statistical grounds. Certainly there is nothing known about the Cui. program that is especially designed to encourage children to avoid attempts to make possibly difficult computations.

Rate of work

In this chapter next to nothing has been said about the comparative rates of work of the two samples, and for two reasons. (a) Little evidence was collected on this aspect of performance. In Table 10 the mean numbers of examples attempted in the Com. Test as a whole were reported as 56.69 for the Cui. sample and 55.19 for the Tra. sample, with S.D.'s of 14.69 and 14.74, respectively. The slight advantage of the Cui. sample is clearly unreliable. As for the sets of skills, it was impossible under the conditions of testing to obtain measures of rate per set, and the only information procurable would have had to be inferred from NAs. Such data were regarded as too ambiguous to be worth much.

(b) At best, rate of work is generally regarded as a questionable criterion of computational proficiency. The most certain way to secure a high rating for speed is to record guessed (and incorrect) answers as rapidly as possible for all examples. Accuracy is a far better criterion of mathematical skill, and for this reason it was made the standard for determining the relative success of the samples, and so, of the two programs.

Critical examples

The last analysis consists in a comparison of the records of the two samples on critical examples. A critical example is arbitrarily defined as one in which one sample surpassed the other by having 12 or more correct answers, 10 per cent of the possibility (there being 120 subjects in each sample). Below are listed in Table 21 all the examples in the Com. Test which satisfy this criterion, together with other relevant information.

Perhaps the single most striking fact discernible in the summary is that so few items were critical as that term is here defined, only eight of the 72 items in the Com. Test. In one set there is no such item. In five other sets there is but one.

Table 21

Examples in which the Samples' Numbers of Correct Answers Differ by 12 or More

Set	Item	Superior sample, with margin of correct answers	Number of NAs	
			Cui.	Tra.
1.	13. 2 x ____ = 12	Cui. by 12	2	6
	19. ____ ÷ 2 = 7	Cui. by 21	25	26
	20. ____ + 8 = 8	Cui. by 29	9	7
2.	24. 18 = 7 + ____ + 6	Cui. by 21	22	38
3.	34. 25 + 37 + 8, or 25 37 8 ____	Tra. by 12	18	11
4.	None			
5.	70. Which of these numbers divide into 12 exactly? Put X on them. 7 4 6 5 3	Cui. by 13	68	80
6.	49. 12 + 26 - 24 = ____	Tra. by 14	27	24
7.	43. 1s.1d. - 8d. = ____s. ____d.	Tra. by 14	26	22

As has happened before in other comparisons, the Cui. sample excelled here in numbers of critical examples in their favor, five of the eight.

Six of the items--19, 24, 70, 49, and 43--should probably be rejected because of the very numerous NAs in one or both samples. Twenty-four NAs represent 20 per cent of the number of subjects in each sample, and this percentage is equalled or exceeded in each of the items named by one or both samples.

Four of the five examples in which the Cui. subjects have the greater number of correct answers are number sentences in which the missing numeral is to be entered in a position other than final. For the Tra. subjects these items scarcely measured achievement, for, as has been stated before, except for a few minutes of pre-test practice, they were unaccustomed to such examples. The fifth example in which the Cui. sample excelled, item 70, has to do with number divisibility, a mathematical concept comparatively strange to most Primary III Tra. children. Hence, this item (and the other two like it) might well have been disapproved by the Tra. panel as inappropriate for the assessment of achievement. In any case the comparison of the accuracy scores for item 70 is meaningless because of the very large frequencies of NAs in both samples.

To turn to the three critical examples in which the Tra. subjects were the more accurate: item 34 did not measure achievement for the Cui. subjects as it did for the Tra. subjects. Many of the Cui. subjects, perhaps more accustomed to the horizontal form of addition examples, showed by their work on test papers that they were confused by the direction to "Add either way," and did all sorts of peculiar things. On the other hand, the Tra. subjects uniformly used the example expressed in vertical form.

Item 49 also introduced factors which militated against its being a measure of achievement for the Cui. subjects. Had this item and others of the same type been written with brackets or parentheses as is common in the Cui. program, the Cui. subjects would almost certainly have done better than they did. As for item 43 (Scottish money) there is no accounting for the apparently greater success of the Tra. subjects; but whatever significance may seem to attach to their greater number of correct answers is wiped out by the fact that this example was omitted by 54% of the Cui. subjects and by 67% of the Tra. subjects.

Concluding Statement

To the extent that the Com. Test provided measures of achievement on the part of both the Cui. and the Tra. samples--and there is every reason to believe it did so with comparatively few exceptions--the Cui. and the Tra. subjects were equally proficient in the computational skills investigated. Or, to state the matter differently, each program was as effective as the other in engendering competence in the skills taught in the two programs. Or, to use still another wording, if arithmetic instruction in computation in the first three years in Scottish schools is limited to the skills represented in the Com. Test, substantially the same results are to be anticipated whether the Cui. or the Tra. program is adopted. Not one of the forms of analysis employed in this chapter turned up anything to contradict this conclusion.[6]

The conclusion, no matter how stated, is warranted only as long as measures of achievement are based upon computational skills taught in the two programs, as here, and it is not to be interpreted as implying that the programs are a match for each other for the totalities of computational skills that may be taught in them. As will be seen in the following chapters, these totalities are quite unlike each other.

---

[6] These statements are made in the present tense as if the Tra. and the Cui. programs taught to the research subjects beginning in 1962 still persist without change in 1968, as they do not. Both programs, during the intervening years, have been steadily modified, a fact made amply clear in the accounts of the programs written by Miss Law and Mr. Adams. See Chapter I.

CHAPTER VI

EXTENT OF TRANSFER: THE TRA. SUBJECTS

In this and the following chapters the focal issue is transfer of learning. The attempt will be made to determine the extent to which the samples of Tra. and Cui. subjects were able, with untaught skills, to utilize the ideas, understandings, and procedures they had acquired in pursuing the particular program they had studied. In this chapter attention will be centered on the Tra. sample as it worked on the Cui. Test. For a copy of the Cui. Test see the Appendix.

This test is comprised of computational items challenged by the Tra. panel when presented by the Cui. panel for inclusion in the Com. Test. Or, rather, it is composed of these items, plus others of like and unlike character added to make the final 35-minute Cui. Test. The assumption was that all items in this test should have been taught to all the Cui. subjects, and none of them to any Tra. subjects.

Actually, this assumption could not be fully satisfied. In several earlier sections of this report mention has been made of diversity of practice in both the Cui. and the Tra. groups of cooperating schools. And, as will be seen, some skills in the Cui. Test had not been taught thoroughly in a few Cui. schools, while certain of these skills had been at least introduced in Tra. schools. To make this admission is to invite the criticism that these violations of the assumed conditions of learning defeats the possibility of distinguishing between achievement and transfer. And this criticism would be sound if close similarity actually prevailed,--as it did not. Rather, there were occasional similarities (but not identities) at different points, and these, to different degrees of depth.

The contents of the foregoing paragraph are intended merely to caution against over-simplified interpretation of the results of comparisons in this and the succeeding chapter. To illustrate, the data for the two samples on the Cui. Test as a whole may be examined.

First, according to Table 1, Chapter III, the Cui. and the Tra. samples (a) were of very nearly the same age, though the subjects in the latter were less homogeneous, (b) had had very like total amounts of arithmetic instruction in the three years preceding the testing, and (c) had mean intelligence scores within 0.03 of each other. As far as these measures are concerned, therefore, the samples were comparable, and possible differences in test scores are not to be accounted for in terms of these three factors.

According to Table 2, the mean scores for accuracy on the Cui. Test are: Cui. sample, 21.78 (S.D., 13.65); Tra. sample, 15.43 (S.D., 8.73), and the mean scores for rate are: Cui. sample, 45.21 (S.D., 14.56); Tra. sample, 40.21 (S.D., 16.01). The differences between means in both Ac. and rate, in favor of the Cui. sample, are significant at the 0.01 level.

If one were to forget for the moment the caution or warning stated above, one could argue that the Tra. sample by earning an Ac. mean 71 percent the size of the Cui. sample mean and a rate mean 88 percent that of the Cui. sample, demonstrated a huge amount of transfer. This inference would of course be based upon the false hypothesis that the Tra. subjects had had no instruction on the skills comprising the Cui. Test.

More dependable evidence on transfer may be forthcoming from comparisons of the records of the two samples on eight segments of the Cui. Test in the following chapter section. These segments are sets of relatively homogenous skills, concerning most of which something is known regarding the instruction offered in the two programs.

The illustration above serves to exemplify the method to be employed in estimating extent of transfer. In this chapter, the scores of the Cui. sample on the Cui. Test are taken as measures of achievement (the skills had supposedly been taught), and with these measures of achievement the scores of the Tra. sample are to be compared, in order to assess amounts of transfer. In the following chapter, the Tra. scores on the Tra. Test afford measures of achievement; the Cui. scores on the same Test, the means of estimating transfer.

Results on Types of Skill

Set 1. One-operation items. Items 1 - 9
Table 22

Set 1 consists of nine items, five in the form of simple number combinations in division and multiplication, four of the types 27 - 6 = _____ and 90 ÷ 3 = _____. All items were selected, it must be remembered, in consonance with the Cui. program. The number combinations (5 x 6 = _____, 8 x 8 = _____, 5 x 10 = _____, 10 x 10 = _____, and 32 ÷ 8 = _____) reflect the Cui. emphasis on 5, 8, and 10 as factors and on doubling and halving. Hence, the Tra. subjects, who regularly made use of tables and in Primary III did not go beyond the so-called 6-tables, were presumably at a disadvantage.

Table 22 presents the gross data for Set 1, which obviously is none too homogeneous in terms of kinds of item. The Cui. Ac. mean, 5.87, represents 65.2 percent, and the Tra. Ac. mean, 5.00, represents 55.5 percent of the possible score.

Table 22

The Cui. Test.
Scores of Samples on One-Operation Examples, Including Five Simple Number Combinations

Scores	Items 1 - 9			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
9	16	0	9	1
8	21	0	12	2
7	17	0	14	1
6	12	2	21	2
5	19	6	14	8
4	15	6	21	9
3	11	12	10	21
2	6	13	5	12
1	1	23	5	14
0	2	58	9	50
Means	5.87	1.26	5.00	1.84
S.D.'s	2.27	1.74	2.49	2.09

As in Chapter V, so here, following the discussion of data in the basic tables comes consideration of the results of item analysis. In Set 1 the Cui. subjects outmatched the Tra. sample in frequencies of correct answers for eight items, the largest margin, 28, being for item 5 (3 x 20 = _____); the next largest margin, 22, for item 8 (4 x 25 = _____), and the next largest, 17, for item 6 (32 ÷ 8 = _____).

The Cui. subjects made the fewer errors for seven items, in two by 10 or more. The average number of errors per example is 24.9 for the Cui. sample and is 30.7 for the Tra. sample. And the average frequencies of NA's are 16.7 for the Cui. and 24.4 for the Tra. samples. Since all subjects in both samples completed Set 1, these NA's represent deliberate omissions.

The results of the comparisons for Set 1 of the Cui. Test are not easy to interpret from the standpoint of transfer of learning. In the matter of accuracy (numbers of correct answers) the Tra. subjects did almost as well as did the Cui. subjects,--85.2 percent as well, and this on items supposedly beyond their ability. Was this because of transfer? Certainly not exclusively so, for the Tra. subjects had had considerably more experience in such multiplications as 4 x 25 than the Cui. subjects had had. (See the Tra. Test.) On this account it is surprising that they were so much less successful with examples of this kind than were the Cui. subjects. The truth is that the Cui. sample did none too well on its own test. On item 6 their accuracy percent was only 52.5, and their percentages were even smaller for items 3, 8, and 9.

In a word, the Tra. sample looked relatively good on Set 1 because the Cui. sample apparently failed to measure up to expectations. It cannot be said with any confidence that the Tra. subjects transferred enough to account for their ability to perform 85.2 percent as well as did the Cui. subjects. The latter subjects were tested on some skills that they do not seem to have learned thoroughly (and the corresponding items might well have been excluded from the Cui. Test), while the Tra. subjects had had the benefit of unanticipated instruction on skills thought to be unknown to them.

Set 2. Fraction computations and two-operation examples with whole numbers and fractions. Items 10 - 23. Table 23.

The name given this set above indicates that Set 2 is hardly homogeneous as to items. Its fourteen items are of three distinguishable types and will be divided into three groups accordingly. Group 2 consists of but three items, and Group 3, of two items. Scores for the three groups are, however, pooled together in Table 23. When the scores for the small groups, as well as for the larger group of nine items, are discussed, the needed data will be supplied.

a. The set as a whole. Neither sample did well on Set 2. (Table 23.) The mean Ac. score of the Cui. subjects, 4.13, is equivalent to only 29.5 percent of the possible score, a very low average for skills which, in the view of the Cui. panel, had been taught. Of course the Cui. mean Ac.

Table 23

The Cui. Test
Scores of Samples on Fraction Computations
and on Two-Operation Examples
with Whole Numbers and Fractions

Scores	Items 10 - 23 Cui. Sample Ac.	 Cui. Sample NA	 Tra. Sample Ac.	 Tra. Sample NA
14	3	2	0	6
13	1	2	0	4
12	4	2	0	3
11	2	1	1	5
10	3	12	1	7
9	5	5	0	7
8	1	3	0	6
7	7	5	1	5
6	5	7	6	6
5	17	6	11	6
4	9	10	18	10
3	13	4	23	5
2	17	9	16	12
1	14	11	13	12
0	19	41	30	26
Means	4.13	3.88	2.54	5.00
S.D.'s	3.66	4.12	2.17	4.51

Note: At the 0.01 level there is a significant difference in favor of the Cui. Group over the Tra. Group in accuracy. At the 0.05 level there is a significant difference in favor of the Tra. Group in numbers of NAs.

percentage is the larger of the two, and reliably so, that of the Tra. sample being 2.54.

The extreme difficulty of the set is shown also by the large number of NA's. For each correct answer the Cui. subjects omitted 0.94 of an example, and the Tra. subjects, 1.98 examples. And in both instances the NA's cited were deliberate omissions, for all subjects in both samples finished Set 2.

Evidently too high a standard of achievement was expected of the Cui. subjects. The Cui. subjects employed in developing the Cui. Test must have been unknowingly more highly selected than were those used in the research testing.

b. Finding fractional parts and writing fractions. Items 10 - 18. The first five items are illustrated by item 10, 1/3 of 12 = _____, item 12, 3/5 of 10 = _____, and item 14, 1/24 of 48 = _____. The sixth is item 15, Find the half of 96. The last three in the group of nine items are, What fraction of 6 is 2? (of 12 is 4?), and (of 20 is 15?).

By the end of Primary III children in the Tra. program had been taught to deal with unit fractions only, and these with denominators no larger than 9.[1] Moreover, such an item as Find $\frac{1}{2}$ of 12 was a computational item involving

---

[1] See Mr. Allan's statement, p. 17 of Chapter I.

"short division" (4/12). Group 1 items therefore presented to them a stiff challenge. As it turned out, they presented a stiff challenge to the Cui. subjects as well. The mean frequency of correct answers per item for the latter was 36.8 (out of 120), while that for the Tra. sample was 25.4. The mean number of errors per example was 53.1 for the Cui. sample and 60.3 for the Tra. sample. And the corresponding NA means were 30.1 (Cui.) and 34.2 (Tra.).

On items 10, 13, 14, and 15 the two samples were substantially equal in accuracy. Each of these items involves a unit fraction (half, fourth, and twenty-fourth), with which, except for the last, at least some of the Tra. subjects were familiar, but only in the sense mentioned above. On the item in which 1/24 appears, but 25 subjects in each sample were successful; and on the item, Find the half of 96, only 39 Cui. and 41 Tra. subjects obtained correct answers.

Work with item 11, 2/3 of 9 = _____, and item 12, 3/5 of 10 = _____, produced 51 and 31 correct answers, respectively, in the Cui. sample compared with 7 and 6 in the Tra. sample.

In item 16, What fraction of 6 is 2?, the totals of correct answers are 16 and 6 (Cui. and Tra.); in item 17, What fraction of 12 is 4?, 14 and 4; and in item 18, What fraction of 20 is 15?, 8 and 4.

c. Two operation examples with fractions. Items 19, 21, and 22. For 1/3 of 9 = 9 ÷ _____, the frequencies of correct answers are 62 (Cui.) and 43 (Tra.). Here the Tra. subjects were able to transfer fairly well their knowledge of 1/3 of 9, perhaps only as $3\overline{)9}$, to a rather new type of number sentence. Forty-three of them secured correct answers, compared with 62 Cui. subjects. On item 21 8/8 of 8 = 0 + _____, and item 22, 1/8 of 16 = ½ of _____, only about a quarter of the Cui. subjects and about an eighth of the Tra. subjects computed successfully. The limited success of the Tra. subjects on the last item--and it is limited--may have been the result of dealing with the unit fractions ½ and 1/8 only; but how they were able to deal with 8/8 is an open question.

As would be expected from the small numbers of correct answers for the examples in Group 2, the frequencies of negative responses are large: a mean of 82 for the Cui. sample and a mean of 95 for the Tra. sample.

d. Two-operation examples with whole numbers. Items 20 and 23. The two items in Group 3 are 8 x 4 = 16 x _____ and 10 ÷ 2 = 4 + _____. As was stated above, these examples, since they contain no fractions, hardly belong to Set 2. Once again here are items too difficult to have been included in the Cui. Test, for the Cui. subjects obtained only

20 and 26 correct answers on them. The Tra. subjects did not do too badly, with successes by 10 and 12 children. If the Cui. subjects had been taught to work this kind of example, they had not learned too well,--not much better than had some Tra. subjects who had had no instruction at all.

To summarize: the Tra. subjects seem to have been able to deal with computational examples involving unit fractions (except 1/24) most probably by using "short division." To do so they could have drawn upon transfer to help them with unfamiliar examples. In the case of items 11 and 12 the Tra. children were unable to cope with 2/3 of 9 and with 3/5 of 10, perhaps never having even seen such fractions, to say nothing about having had no instruction on the method of computation. For these items they had little to transfer, as they did also in the case of the (to them) totally strange requirements of items 16 - 18; e.g., What fraction of 12 is 4?

Set 3. Computation with one and two brackets, fractions and whole numbers.
Items 24 - 30. Table 24

Set 3 consists of seven items. The first two items contain a single bracket each; the last five, two. Three contain only whole numbers; four, fractions in one or both brackets. Illustrations of the more complicated sort are: [40 + 4] - [10 + 4] = _____ and [4/5 of 10] -[$\frac{1}{2}$ of 6] = _____.

The reader may recall that, not having had any previous experience with examples involving brackets, the Tra. subjects, with the consent of the Cui. panel, were given a pretest period of perhaps 15 minutes in working simple examples of this kind. The purpose was to eliminate to some degree one factor of difficulty extraneous to the measurement of computational skill in this study, thus making it possible to concentrate on the computations called for.

The data in Table 24 show that the Ac. mean of the Cui. sample, 2.77, is understandably 0.91 larger than that for the Tra. sample. In terms of percentage the two means are equivalent to 39.6 and 26.6.

Only two subjects in the Cui. sample and one in the Tra. sample did not finish even the first example in Set. 3 within the time limits. On the average the Cui. subjects had 0.36 fewer omissions per example.

The Cui. subjects surpassed the Tra. subjects in accuracy in six of the seven examples by securing larger totals of 15 or more correct answers in each. Yet, the Cui. subjects had their difficulties. On four items (26, 28, 29, 30) 50 or fewer of the 120 members of this sample obtained correct answers. In other words, on more than half of the items the Cui. sample had accuracy percentages of 42 or less.

Table 24

The Cui. Test
Scores of Samples on Examples with One and with Two Brackets, Fractions and Whole Numbers

Scores	Items 24 - 30			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
7	8	11	0	3
6	8	3	2	6
5	8	3	3	12
4	17	6	15	14
3	22	16	18	18
2	18	15	29	4
1	18	9	24	17
0	21	57	29	46
Means	2.77	1.74	1.86	2.10
S.D.'s	2.09	2.37	1.52	2.16

The Tra. subjects found item 24, 4 + [$\frac{1}{2}$ of 8] = _____, the easiest of the seven; 90 of them, compared with 84 of the Cui. subjects, had correct answers. On item 25, [6 x 3] ÷ 9 = _____, the Tra. subjects had 47 correct answers (Cui., 66); on item 27, [2 x 4] - [4 x 2] = _____, 46 correct answers (Cui., 71); and on item 26, [40 + 4] - [10 + 4] = _____, 27 (Cui., 50). On each of the remaining three items, 28 - 30, the Tra. subjects were successful in making the required computations in but 11 or fewer attempts.

In conclusion, it is to be noted that the three items just mentioned as the most difficult for the Tra. sample called for work (a) with two brackets and (b) with such fractional expressions as 3/4 of 8, 4/5 of 10, and 3/4 of 16. Their program had equipped the Tra. subjects for neither one of these sources of trouble, either one of which would have occasioned considerable difficulty. Moreover, they had little in their arithmetical backgrounds to assist them through transfer. However, their record with the first four items, three with whole numbers and one with ½ of 8, and two with two brackets,--their record with these items is another matter. To obtain correct answers for them amounting to from 27 to 90 on each, they must have been able to benefit a great deal from transfer; that is, unless, as is unlikely, it can be presumed that pretest practice with brackets provided enough learning in itself to account for their special success with these examples.

Set 4. Number meanings; doubling numbers. Items 31 - 35. Table 25.

Set 4 calls for two instances of doubling numbers (item 31; Double 48 and item 35, Double 250.) and for one item, 32, related to doubling (100 - 50 = _____), as well as for two items which can be subsumed under the heading, number meanings (33. What number comes after 99?_____; and 34. Put X on the 4 that means 4 tens. 4 4 4).

The data in Table 25 indicate little difference in the ability of the two samples to find answers for the five items in Set 4 as a whole. The pairs of means both for Ac. and for NAs are substantially the same. Moreover, they are directly comparable since all but one Cui. subject and all but four Tra. subjects completed the set.

Table 25

The Cui. Test
Scores of Samples on Number Meanings
Doubling Numbers

Scores	Items 31 - 35			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
5	23	7	25	11
4	28	3	30	1
3	24	6	18	8
2	24	9	20	9
1	11	32	13	22
0	10	63	14	69
Means	2.98	0.96	2.93	1.03

On the three doubling items the Cui. subjects secured an average of 13.7 more correct answers than did the Tra. subjects. On the two items designated as involving number meanings the advantage in accuracy lay now with one sample, and now with the other. The Tra. subjects had 6 fewer correct answers for item 33 and 21 more correct answers for item 34.

On the whole, the performances of the two samples shed no light on the problem of transfer. The Tra. subjects did as well as they did on Set 4, not because they successfully carried over previous learning to new skills or applied it in new contexts, but because the Tra. program provides for instruction in doubling numbers, if less than does the Cui. program, and stresses number meanings more than does the latter program. For both samples items 31 - 35 measured the results of direct teaching. Accordingly, in view of the purpose of the Cui. Test which was to be a specialized instrument to measure the achievement of the Cui. subjects alone, it seems to have been a mistake on the part of the Cui. panel to have included in the test at least some of the items in Set 4.

Set 5. Number progressions; relative size of numbers
Items 36 - 60. Table 26.

Of the five items in this set, three call for the completion of the series 5, 10, 20, 40 _____; 20, 18, 16, 14, _____, and 4, 8, _____, 32 _____, 128. Two call for the writing of five given numerals in order of size: 1, 5, 3, 6, 4, and 10, 60, 30, 100, 80.

Both the Ac. means in Table 26 are small; they are equivalent to accuracy percentages of 32.6 for the Cui. subjects and 24.6 for the Tra. subjects. And both samples omitted a good many examples, the Cui. subjects a mean of 27.2 for the first five items; the Tra. sample, a mean of 35.5. Since 113 Cui. and 111 Tra. subjects worked through Set 5, practically all the omissions were the result of intent.

The Cui. subjects have the greater number of correct answers for each of the first three items (progressions); the differences in their favor range from 2 to 15; but the frequencies of negative responses on all these items are large for both samples: Cui., 81 to 101; Tra., 85 to 113. The Ac. means expressed as percentages are: Cui., 21.4; Tra., 12.8.

Both samples improved their records on the last two items (number meanings), and the Cui. sample excelled in accuracy in both by margins of 9 and 11 correct answers.

To interpret: Evidently the instruction assumedly given the Cui. subjects on progressions had not been very fruitful. The Tra. subjects did fairly well on the first three items, and did so with little or no teaching of progressions. Such success as they had must be attributed, as far as is known, to transfer from the skills of counting by 5's and from kindred activities.

Transfer, however, is not the explanation for the comparatively good showing of the Tra. subjects on number meanings (items 59 and 60). In their program, as in the Cui. program, children are taught the skill of arranging numerals according to size. Possibly what was done in the Cui. program was somewhat more thorough and extensive.

Table 26

The Cui. Test
Scores of Samples on Number Progressions and on Comparative Size of Numerals

Scores	Items 36 - 40			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
5	8	9	1	12
4	6	2	2	4
3	16	11	18	16
2	31	19	28	15
1	22	13	24	23
0	37	66	47	50
Means	1.63	1.14	1.23	1.48

Set 6. Two- and three-operation examples with one or two sets of brackets. Items 41 - 47. Table 27.

The last three items in this set of seven are illustrative: [3/4 of 100] - 5 = _____; [4 x 5] ÷ [5 x 4] = _____, and [5 + 6] x [14 - 12] = _____. Four items contain only whole numbers; the other three involve the fractional expressions $\frac{1}{2}$ of _____ (actually $\frac{1}{2}$ of 6), 1/3 of 18, and 3/4 of 100. All computations are clearly characteristic of the Cui. program and absent in the Tra. program.

Both Ac. means in Table 27 are small (not surprisingly so in the case of the Tra. sample) and amount to percentages of 19.6 for the Cui. subjects and 9.7 for the Tra. subjects. The totals of negative responses are correspondingly large, ranging from 90 to 100 per example for the former and from 100 to 118 for the latter.[2]

The number of Cui. subjects securing correct answers is greater than the corresponding number of Tra. subjects for all seven examples by differences of from 7 to 17, amounts that in the circumstances are not large. The ranges of correct answers per example are: for the Cui. subjects 10 to 30 on items 45 and 44, respectively; for the Tra. subjects, 12 to 20, respectively, on the same two items.

To sum up: Set 6, with an average percentage of 19.6 in accuracy, seems to have been too difficult for the Cui. subjects, to have been included in a test designed to measure achievement in skills supposed to have been thoroughly taught. While their average percentage for accuracy, 9.7, is admittedly low, the Tra. subjects did about half as well as the Cui. subjects. The examples are complicated and do not appear in the Tra. program; the significance of brackets was totally unfamiliar to the Tra. subjects prior to the brief pre-test period of practice; and the presence of fractions in strange settings did not help them a bit. For whatever they were able to accomplish on Set 6, credit appears to be due to the transference of general ideas relating to computational procedures.

---

[2] Thirteen Cui. and 15 Tra. subjects did not have time to attempt Set 6, and the two samples are charged with these numbers of NAs for each example. If these NAs are omitted, the Cui. range of negative responses is 77 to 87, and the Tra. range, 85 to 108. Obviously, these totals are still high. Every total represents more than 50 per cent of each sample. Moreover, the relationship between the samples with respect to negative responses is altered very little.

Table 27

The Cui. Test
Scores of Samples on Two- and Three-Operations Examples
with One or Two Sets of Brackets

Scores	Items 41 - 47			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
7	3	18	0	19
6	2	7	0	9
5	7	12	4	14
4	9	9	3	11
3	10	12	2	7
2	6	10	9	16
1	18	9	25	16
0	65	43	77	28
Means	1.37	2.74	0.68	3.08
S.D.'s	1.94	2.65	1.20	2.57

Set 7. Computation with Scottish money, all with fractions. Items 48 - 53. Table 28.

All six of the examples in Set 7 have to do with Scottish money, and all six involve fractions in one way or another. Sample items are: 5/12 s. + $\frac{1}{4}$s. = _____ pennies; £3/4 = £$\frac{\underline{\quad\quad}}{20}$; What fraction of a pound is 2s.?_____.

The Tra. program, like the Cui. program, gives a good deal of attention to money, but not in the complicated forms and not in the fractional relationships to be found in Set 7. It is therefore small wonder that the Tra. Ac. mean is only 1.04 (Table 28), or in percentage only 17.3. On the other hand, the Cui. Ac. mean is but 1.51, or 25.2 percent. Only on item 48, $\frac{1}{2}$ of ____ = 4d., did as many as 62.5 percent of the Cui. sample have correct answers. On item 53, 3s. + 3s. + $\frac{1}{2}$ of £1, 42.5 percent of the Cui. subjects had correct answers. On no other item did the percentage reach 20.0. While generally making fewer correct computations per example, the Tra. subjects were never at a greater disadvantage in this respect than 14. Interestingly enough, the Cui. subjects have the greater frequency of errors on every example (from 3 to 9)--evidence perhaps that they recognized the examples as those they should be able to work. The Tra. subjects have the greater number of NAs for each example, in frequencies from 14 to 25.

Table 28

The Cui. Test.
Scores of Samples on Examples with Scottish Money,
All with Fractions.

Scores	Items 48 - 53			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
6	3	20	0	35
5	6	14	2	12
4	4	11	1	18
3	11	17	8	13
2	27	7	30	9
1	30	15	27	8
0	39	36	52	25
Means	1.51	2.62	1.04	3.39

The statements above are made without regard for the fact that 15 Cui. and 27 Tra. subjects did not have time to work on Set 7. If the NAs resulting from lack of time are excluded, the range of negative responses per example for the Cui. sample is from 15 (item 48) to 86 (item 52), and for the Tra. sample is from 42 to 90 on the same two examples (90 also on items 49 and 51). And the Ac. means adjusted for the same reductions of subjects are: Cui., 1.73; Tra., 1.13.

On the whole, the Cui. program does not appear to have been effective in developing the skills called for in Set 7, and the Tra. program, which taught these skills to a minimal extent, did inculcate enough understanding for the Tra. subjects to profit somewhat from transfer.

Set 8. Measurement examples, plus two miscellaneous items. Items 54 - 60. Table 29.

Five of the seven items require knowledge of linear, liquid, or avoirdupois units and skill in exchanging units. Two items, the last two in the set, do not properly belong in the set: How much must you add to 24 to get 63? and How much greater is $\frac{1}{2}$ of 24 than $\frac{1}{4}$ of 16? They will be discussed separately below.

To refer at first to the data in Table 29: it is obvious that Set 8 was overly difficult for the purposes of this research. The Cui. Ac. mean is small; the Cui. NA mean is large, and only 12 Cui. subjects made total scores of 5 to 7 while 69 made scores of 0 or 1. The Cui. Ac. mean in percentage is but 23.9. If the 26 Cui. subjects who did not have time for Set 8 are excluded, the Ac. mean in percentage is only 30.4, to be compared with 22.9 for the Tra. subjects under the same conditions. Moreover, the numbers of negative responses made by the Cui. sample range from 66 to 108 per example on the measurement items, equivalent to

Table 29

The Cui. Test.
Scores of Samples with Examples Involving Measurement,
Plus Two Miscellaneous Examples

Scores	Items 54 - 60			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
7	3	26	0	32
6	2	5	1	8
5	7	3	2	12
4	11	10	7	12
3	14	15	18	11
2	14	10	19	14
1	18	23	16	13
0	51	28	57	18
Means	1.67	2.96	1.27	3.79
S.D.'s	1.92	2.65	1.48	2.58

55 percent or more of the subjects in each instance. Seemingly, the Cui. panel over-estimated the degree of achievement to be expected from the Cui. subjects.

On four of the five measurement items the Cui. subjects were more accurate than the Tra. subjects; on one, the Tra. subjects excelled; but all differences between frequencies of correct computations are small--9 or less. Furthermore, the differences between the frequencies of negative responses for the two samples amount to 9 or less, the Cui.

subjects having the edge on four items.

On both of the last two items, in which measurement plays no part, the Cui. subjects have the larger frequencies of correct answers by 6 and 7; but this comparison means little, for those of the Cui. subjects numbered but 34, and those of the Tra. subjects but 20. With this statement, the results for these two items may be dismissed from further consideration.

Finally, as has been true more than once before, it is impossible with confidence to determine the significance of the data for the measurement items with respect to transfer of learning. The issue is clouded by at least two factors. In the first place, the achievement of the Cui. subjects, which provides the basis for interpretation, is so poor that the even poorer record of the Tra. sample is made to look relatively good. From this circumstance a considerable amount of transfer on the part of the Tra. subjects might be inferred.

In the second place, the hypothesis implicit in the foregoing sentences is certainly unsound,--the hypothesis that the Tra. subjects had received no instruction on skills involving measurement. On the contrary, the Tra. program provides for the teaching of some aspects of measurement, though the extent is not known. It is safe to say that in all Tra. schools children acquire the concepts represented in measurement units, but probably in few classes do they learn to use them in the kinds of computation called for in Set 8. But it would not take many such classes to account for the Tra. showing on the set.[3]

In the end, then, one can scarcely afford to take a stand on the question whether the Tra. sample profited much or little from transfer in dealing with the measurement items of Set 8.

Concluding Statement

In this chapter comparisons have been made between the records of the Cui. and the Tra. samples on eight sets of items in the Cui. Test. Not all of the sets are as homogeneous as could have been desired. For the Cui. subjects the Cui. Test measured achievement, since the test items call for skills which, in the judgment of the Cui. panel, had been taught to them. For the Tra. subjects the Cui. Test was intended to furnish evidence on the extent to which they were able to transfer learning to skills which, by hypothesis, they had not been taught.

---

[3] One could hypothesize that the Tra. schools had taught as much about computational skills with measurement units as had the Cui. schools, in which case the comparisons for Set 8 would be of achievement with achievement, and the problem of transfer would not arise. This possibility can be dismissed as exceedingly unlikely. Note the simple kinds of measurement items approved by the Tra. panel for use in the Com. Test.

Unfortunately, the purpose of the comparisons could not be realized any too well. Examination of performances (correct answers, errors, and omissions) revealed two complicating factors. (a) In some instances the Cui. panel in constructing the test anticipated from the Cui. subjects a much higher level of achievement than eventuated. The best explanation seems to be that the Cui. classes used in standardizing the test were unwittingly more highly selected than were those employed in the research proper. (b) And again in some instances, the Tra. sample, or, better, some of the Tra. classes, were known or were strongly suspected of having had instruction on supposedly untaught skills.

The effect of condition (a) was to make the showing of the Tra. sample look better than it probably was. For example, when the Cui. Ac. means are small, though the Tra. means are smaller, one is tempted to infer that the Tra. sample profited from transfer more than it may have.

The effect of condition (b) was to make the study of relative performances comparisons of achievement with achievement, thus destroying the chance to find evidence of transfer.

Below are listed the results of comparisons set by set, together with brief comments.

Set 1. One-operation items, nine items. Five of the items are simple number combinations. Previous instruction given the Tra. sample and a relatively poor showing of the Cui. sample combine to make uncertain the amount of transfer on the part of the Tra. subjects.

Set 2. Fraction computations and two-operation examples with fractions and whole numbers; 14 items. The Tra. sample seems to have transferred to good effect their understanding of unit fractions in computation, by converting them to "short division." Other skills required offered little chance for transfer.

Set 3. Computation with one and two brackets, fractions and whole numbers; seven items. Tra. subjects seem to have transferred learning to a considerable extent in the two one-bracket examples and in two two-bracket examples, the latter with small numbers and simple computations.

Set 4. Number meanings; doubling of numbers; five items. Both samples had been taught the skills called for; hence, evidence of extent of transfer is unavailable.

Set 5. Number progressions; comparative size of numerals; five items. The Tra. subjects did rather well with number progressions (not taught), by transferring from such activities as counting by 5's. Items involving the arrangement of given numerals according to size had been taught to both samples.

Set 6. Two- and three-operation examples with one or two sets of brackets; seven items. Cui. subjects did poorly on the set though supposedly taught the necessary skills; Tra. subjects, with only the brief pre-test practice period, did half as well. To do so, they must seemingly have had to transfer general ideas concerning computational procedures.

Set 7. Computation with Scottish money; six items. Cui. classes, at least in some instances, had not generally acquired the skills tested. Tra. subjects, with no direct teaching of these skills, by means of transfer did 65 percent as well as did the Cui. subjects.

Set 8. Measurement examples; five items, plus two miscellaneous items which are disregarded here. The measurement computations were too difficult for the Cui. subjects who were being tested on skills assumed to have been taught (adjusted Cui. Ac. mean, 30.4 percent). No estimate of the extent of transfer on the part of the Tra. subjects is possible, for they all had had instruction on Scottish money and possibly some had learned how to make the computations called for.

- - - - - - - -

So much for a summary of the results of the inquiry built around the Cui. Test; left to Chapter VII are theoretical problems concerning the transfer of learning. The reader may have noted that at no place in this chapter has any such statement been made as, "The Tra. subjects transferred none of their previous learning."

CHAPTER VII

EXTENT OF TRANSFERS: THE CUI. SUBJECTS

In this chapter the purpose is to determine the extent to which the Cui. sample was able to transfer understandings, concepts, and skills taught them in the Cui. program to successful performance on a computational test for which their program, by assumption, had not been prepared to measure their achievement. Chapter VII is therefore the counterpart of Chapter VI, in which the computational skills of the Tra. sample were assessed by the Cui. Test, to ascertain extent of transfer.

An examination of the Tra. Test (for a copy see the Appendix) reveals no items of certain types which are characteristic of the Cui. Test,--and this dissimilarity was of course intended. The Tra. panel used kinds of item rejected by the Cui. panel in the preparation of the Com. Test, extended the skills in these items to make more difficult items, and added others thought not to be taught in the Cui. program. In the Tra. Test there are no items involving fractions, or number sentences with brackets, or Scottish money, or linear, avoidupois, and volume measures. Items in the first- and last-named categories in the Com. Test represent the extent of coverage of instruction on these topics in the Tra. program. All except nine of the items in the Tra. Test call for straight-forward computation with abstract numbers in the four operations. The exceptions were intended to probe deeper than the related items on the Com. Test into understanding of the meaning and structure of the decimal system of number notation.

The mean accuracy scores of the two samples on the Tra. Test are: Cui., 26.22; Tra., 32.26, with S.D.'s of 16.75 and 15.72, respectively (Table 10). The mean scores for rate of work are: Cui., 45.83; Tra., 46.77, with S.D.'s of 13.84 and 13.13, respectively. The Tra. subjects were understandably the more successful (reliably so) in securing correct answers, if only slightly so (and unreliably) in the numbers of items attempted within the 35-minute time limit. In terms of percentages of correct answers the Cui. sample on their achievement test (the Cui. Test) outscored the Tra. sample on the same test (a transfer test for them) by 8.4; and the Tra. subjects on their achievement test (the Tra. Test) outscored the Cui. subjects on the same test (for them, a transfer test) by 9.6.

These comparisons would seem to indicate that the Tra. program produced more transfer than did the Cui. program. But such comparisons, based as they are on scores on the tests as wholes, can oversimplify the matter and may conceal much of importance. More revealing should be a study of the relative success of the samples on types of item in the Tra. Test, which may disclose, as the gross comparisons cannot, the particular ways in which transfer operated to the advantage of the Cui. sample. Accordingly, the 63 items in the Tra. test were divided into seven sets. Within each set the items are as homogeneous as they could be made. The sets, in order, comprise the following numbers of items: 11, 4, 12, 15, 4, 9, 8. Conclusions concerning relative achievement and transfer for the two 4-item sets are clearly precarious. Moreover, since three of the larger sets include two groups of items that are somewhat different in character, again inferences must be drawn cautiously.

Set 1. Column addition.
Items 1 - 6, 32 - 34, 46, and 60. Table 30

a. The set as a whole. Columns (4) and (7). Set 1 consists of 11 examples in column addition. In six there are four addends of one- and two-place numerals; in three, three addends of two- and three-place numerals; in one, four addends of two- and three-place numerals, and in one, four addends of one-, two-, and three-place numerals.

Columns (4) and (7) of Table 30 show the distributions of scores for the two samples on the set as a whole. The difference between the Ac. means is 1.11 in favor of the Tra. subjects and is reliable. The difference between NA

Table 30

The Tra. Test. Scores of Samples in Column Addition

Scores	Cui. Sample						Tra. Sample					
	Items 1 - 6		Items 32 - 34, 46, 60		Total Preceding		Items 1 - 6		Items 32 - 34, 46, 60		Total Preceding	
	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA
(1)	(2)		(3)		(4)		(5)		(6)		(7)	
11					8	2					19	2
10					17	1					24	1
9					14	1					30	0
8					22	1					9	2
7					11	1					6	0
6	39	2			8	1	54	2			7	0
5	26	1	15	39	8	9	33	1	37	11	5	7
4	21	2	26	34	5	4	9	0	22	5	2	5
3	9	1	20	26	7	5	7	2	27	2	3	4
2	15	2	18	3	12	23	7	0	7	11	4	11
1	6	6	5	5	4	35	2	3	8	38	3	36
0	4	106	36	13	4	37	8	112	19	53	8	52
Means	4.29	0.32	2.33	3.50	6.62	1.83	4.68	0.22	3.13	1.18	7.73	1.44
S.D.'s.	--	--	--	--	3.15	2.31	--	--	--	--	3.28	2.23

The difference in favor of the Tra. group over the Cui. group for accuracy is the only significant difference at the 0.01 level.

means, 0.39, in favor of the Cui. sample, is unreliable. Seventy-three of the Tra. subjects and 39 of the Cui. subjects earned scores of 9 - 11, while 20 of the Cui. subjects and 15 of the Tra. subjects have scores of 0 - 2.

Typically, the Tra. program does much more with column addition than does the Cui. program, especially with three-place addends. From the statement on page 13 of Chapter I, one infers that children in many Cui. schools are taught, in the first three grades, no more than to add two-place numerals with sums to 100. (Note that the Cui. Test has no examples in column addition.) Under these conditions the margin of advantage enjoyed by the Tra. subjects in Set 1 was to have been expected, but in amount far beyond the actual 10 per cent. The small size of this difference can be accounted for as the effect of a huge amount of transfer of learning on the part of the Cui. subjects. Either that, or else in some Cui. classes teachers taught column addition much beyond the supposed limits mentioned above. The fact that eight Cui. subjects made perfect scores of 11, that 17 more made scores of 10, and that 14 more made scores of 9 may properly incline one, if tentatively, to accept the second explanation.

b. Group 1. Items 1 - 6. Columns (2) and (5). According to the data in columns (2) and (5) the Tra. sample's mean for accuracy (average number of correct answers per example) is but 0.39 the larger of the two, and its mean for NAs is but 0.10 the smaller of the two. For Group 1, then, the performance of the two samples were about on a par.

On items 1, 2, and 4 (written horizontally, 26 + 30 + 12 + 31; 30 + 4 + 21 + 13; and 20 + 43 + 32 + 94) the two samples were equally accurate. In the first two, the sums of ones are 9 and 8. Hence, no renaming is required. Moreover, the totals are less than 100. In the third item the sum of tens is 18, and the total is a three-place number. While examples of this type may not have been taught to the Cui. subjects, correct computation should have been easy for them.

In items 3, 5, and 6 the sums of one are 26, 18, and 29, and the sums of tens, after renaming the sums of ones, are 15, 24, and 25, to give three-place answers for all three. They therefore could have presented real difficulties to the Cui. subjects on the assumption that the necessary skills had not been taught them. In these examples the Cui. subjects wrote fewer correct answers by 15 to 27 and made more errors by 12 to 22 than did the Tra. subjects. Nevertheless, the Cui. subjects recorded correct answers to the extent of 50 percent or more on each example.

c. Group 2. Items 32 - 34, 46, and 60. Columns (3) and (6). The additions in items 32 - 34 involve three addends. In two, all addends are three-place numerals; in one, there are two addends of this kind and one two-place numeral. In all, the sums in tens' and hundreds' columns must be renamed; in two, the sums in all columns. For each example the Cui. subjects had the fewer correct answers by 10 or more and also had the greater numbers of NAs by 10 or more also.

Twenty Cui. and 15 Tra. subjects did not have time to attempt item 46 (493 + 27 + 854 + 75), and 63 Cui. and 57 Tra. subjects did not reach item 60 (508 + 5 + 80 + 416). Of those who did attempt these two examples the Tra. subjects have the larger number of correct answers by 20 and 17. Erroneous answers were written to about the same extent by the subjects in both samples.

For group 2 of Set 1 the Tra. Ac. mean is 0.80 larger, and the Tra. NA mean is 2.32 smaller than the corresponding means of the Cui. sample. While the Tra. record is much the better, the Cui. record is not at all bad, whether because of transfer or because of "extra" instruction on column addition; that is, teaching beyond the "limits" of the "standard" Cui. program.

To term the Cui. program Miss Law describes in Chapter I "standard" is actually to use a misnomer. The word implies that she, perhaps with others, had exercised delegated authority to prescribe just what should, and what should not be taught in all Primary I, II, and III Cui. classes. Of course nothing of this kind happened, and Miss Law emphasizes in her statement the great diversity of practice in Cui. schools. The most she could hope to do was to depict something like typical practice. This fact is evident in the following quotation of her own words: ". . . by the end of Primary III many teachers expect . . ." Note: many Cui. teachers; not all Cui. teachers.

The only justification for using the word "standard" and other terms like "limits set" and "extra instruction" is that no better method could be found to indicate arithmetical objectives commonly accepted in Cui. schools. It is hoped that with this explanation the reader will understand such expressions in the manner intended. He will be reminded to do so.

To conclude: it is perhaps time to examine whatever evidence is to be had on the question of whether transfer of learning or "extra instruction" (see above) is to be credited with the success of the Cui. subjects in column addition. This will be done for Set 1 as a whole. The evidence alluded to is rather meager.

The records of the five Cui. classes in the sample that had the highest average accuracy scores for Set 1 were examined. There were 35 children in these classes (29 percent of the total sample), and their Ac. mean is 9.06. This mean is 1.37 times the size of the Ac. mean of 6.62 for the whole Cui. sample and is 1.15 times the Ac. mean of the Tra. subjects, all or most of whom should presumably have been taught the necessary skills for the Set 1 items. Moreover, if the scores of these 35 children are eliminated, the Ac.

mean of the remaining 85 Cui. subjects is reduced to 5.62.

True, the Ac. mean of these children on the Tra. Test as a whole, 37.10, is substantially greater than that of the entire Cui. sample, which is 26.22. They were therefore possessed of greatly superior over-all ability in arithmetic. This superiority can be cited to support the view that they were capable of transfer sufficient in amount to explain their high scores on untaught forms of column addition. On the other hand, their superiority, it can also be argued, encouraged their teachers to carry them further in column addition than is suggested in the Cui. program described in Chapter I. And if this "extra teaching" occurred in the five classes mentioned, it probably occurred, though not to the same extent, in other Cui. classes. This second explanation puts less of a strain on credibility, it is believed, than does the first.

Set 2. Horizontal addition.
Items 23 - 26. Table 31

Set 2 comprises but four items, the easiest of which proved to be 5 + 4 + 7 + 6 = _____. The other three examples in the set have five addends, all digits. In two, 0 is an addend, and one of these turned out to be the most difficult for both samples, 8 + 7 + 0 + 8 + 9.

In Table 31, the Ac. means are shown to be 3.05 (Cui.) and 3.28 (Tra.), with S.D.'s of 0.43 and 0.24, respectively. The distributions of scores for the two samples on the set

Table 31

The Tra. Test.
Scores of the Samples in Horizontal Addition

Scores	Items 23 - 26 Cui. Sample Ac.	 NA	 Tra. Sample Ac.	 NA
4	64	8	70	4
3	25	2	29	0
2	14	5	12	4
1	7	4	2	5
0	10	101	7	107
Means	3.05	0.43	3.28	0.24

are very similar. The data as a whole imply that the samples were about equally proficient in the skill in question. Support for this judgment is found in evidence with respect to NAs. All subjects had time to complete the four examples. Such omissions as there were are therefore viewed as having been made deliberately, and they were not too numerous. The largest numbers of NAs are 17 and 15 for the Cui. subjects, and 8 and 9 for the Tra. subjects. All differences between totals of NAs, example by example, are 8 or less.

In two examples, 24 and 26, the Tra. subjects had the greater number of correct answers, by 10 and 13. Incorrect answers were numerous in each sample for the two items, but the differences in frequencies of errors are all 5 or less.

Finally, the results of the comparisons respecting Set 2 require interpretation. The first significant fact to be noted is that the Com. Test contains one item in horizontal addition, 4 + 5 + 6 = _____. That this item was approved by the Cui. panel for the Com. Test means that horizontal addition was taught in Cui. schools, at least to the extent of adding three digits; and there is the possibility that, as in the case of column addition, some Cui. classes carried the skill much further. At any rate, the Cui. subjects in dealing with the Set 2 items were called upon merely to extend a known skill a little. If this extension be called transfer of learning, it was relatively slight.

As a matter of fact, Set 2 might well have been omitted from the Tra. Test as a test of achievement in skills peculiar to the Tra. program. It is possible that the Tra. panel in making up the test overlooked the presence of horizontal addition in the Com. Test, or at least failed to consider the implications.

Set 3. Subtraction.
Items 2 - 7, 35 - 37, 43, 58, 63. Table 32.

a. The set as a whole. Columns (4) and (7). Skill in subtraction was measured by means of 12 examples, which are divided below into two groups. In Table 32 the Ac. mean of the Tra. sample, 6.85, is seen to be larger by 1.64 than the Cui. Ac. mean. The difference between the means is reliable. Also, the Tra. subjects on the average omitted the fewer items by 0.39; but the means of both samples for NAs are large,--more than 2.0.

The distributions of scores for the two samples are quite dissimilar. Fifty-four Tra. subjects and 33 Cui. subjects made scores of 9 - 12; and 32 Tra. and 46 Cui. subjects, scores of 0 - 3.

On the set as a whole the Tra. sample demonstrated clear superiority over the Cui. sample in subtraction as that skill was measured in Set 3. That the Tra. subjects should have excelled is no more than should have been anticipated, for subtraction of the types represented in the set are definitely prescribed by the Tra. program and not by the Cui. program.

b. Group 1. Items 7 - 12. Columns (2) and (5). The items in this group are similar in that in each a two-place numeral is to be subtracted. They are dissimilar in other ways. In three there are two-place sums; in three, three-place sums. In two, renaming of numerals is required after each of two subtractions. According to the results, the easiest and the most difficult items for both samples were, in horizontal form, 79 - 35 and 400 - 35, respectively. All subjects were able to attempt the whole group of examples within the time limits.

Table 32

The Tra. Test. Scores of the Samples in Subtraction

Scores	Cui. Sample						Tra. Sample					
	Items 7 - 12		Items 35 - 37 43,58,63		Total Preceding		Items 7 - 12		Items 35 - 37 43,58,63		Total Preceding	
	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA
(1)	(2)		(3)		(4)		(5)		(6)		(7)	
12					6	3					12	2
11					8	0					16	0
10					9	1					15	0
9					10	3					11	2
8					5	1					8	0
7					8	2					9	4
6	22	3	11	16	10	9	44	4	17	13	9	9
5	18	1	13	4	6	3	23	0	16	6	5	3
4	21	1	10	4	12	4	15	0	17	2	3	2
3	13	3	13	9	4	8	6	0	17	5	4	5
2	9	3	6	40	8	39	7	3	10	33	6	33
1	25	3	12	5	23	6	17	5	8	14	14	13
0	12	106	55	42	11	41	8	108	35	47	8	47
Means	3.23	0.38	1.95	2.03	5.21	2.40	4.07	0.29	2.74	1.76	1.85	2.01
S.D.'s	--	--	--	--	3.88	2.82	--	--	--	--	4.00	2.57

The difference in favor of the Tra. group over the Cui. group for accuracy is the only significant difference at the 0.01 level.

The Tra. subjects have the larger mean for accuracy (4.07 compared with 3.23) and the smaller mean for NAs (0.29 compared with 0.38). For each example the Tra. sample has the larger number of correct answers, by 10 or more in four; the smaller number of errors, by 14 or more in three, and the smaller number of NAs, by 6 or less. While the number of examples is too small to make meaningful any measure of the reliability of differences, the record of the Tra. subjects on the six examples used is consistently the better of the two.

In Cui. schools where the teaching of subtraction was restricted to examples with sums to 100, the children so taught (".....all but..... [the] less able") were prepared to perform the computations in examples 7 - 9; but the Cui. subjects tested did as well as did the Tra. subjects on only one, 79 - 35. On the other two examples the Cui. subjects wrote fewer correct answers by 10 for 94 - 69 and by 17 for 86 - 49.

Presumably, instruction had not made the Cui. subjects capable of working the examples, 135 - 29, 115 - 68, and 400 - 25. Yet, their average for correct answers on these examples represents 43 percent. Success like this implies a large amount of "extra" teaching or a large amount of transfer or, since the dichotomy in the two preceding explanations is questionable, a combination of transfer and "extra" teaching.

c. Group 2. Items 35 - 37, 43, 48, 63. Columns (3) and (6). In terms of the averages reported in Table 32, the Tra. subjects were the more accurate to the extent of 0.79 of a correct answer and omitted fewer examples to the extent of 0.27. Thirty-three Tra. and 24 Cui. subjects secured scores of 5 and 6, and 43 Tra. and 67 Cui. subjects, scores of 0 and 1. Without statistical evidence of reliability (inappropriate with a scale of but 0 - 6), it is still possible to credit the Tra. subjects with a performance much superior to that of the Cui. subjects on the six items of group 2; and there is ample reason to believe that this is true.

The Tra. subjects have 13 or more correct answers and made fewer errors in every example,--10 to 22 fewer errors in four examples. The Tra. sample also omitted fewer items in five instances, in three by 7 to 9.

All the items in group 2 came in the last third of the test. There were therefore many NAs caused by time limitations,--from 10 and 12 for the Cui. and the Tra. samples, respectively, for items 35 - 37, to 77 and 71 for item 63. The totals of omissions owing to shortage of time are so similar for the two samples at successive points in the test that the comparisons in the immediately foregoing paragraph are warranted.

Of the Cui. subjects who did attempt to get answers for the six examples, 41 percent to 56 percent were successful in each. The numbers of such individuals are small in the case of the last two examples,--51 for item 58 and 43 for item 63. For items 35 - 37 and 43 the numbers are quite respectable, 90 or more in each case.

Interpretation of the results of the comparisons above is difficult if one seeks an answer to the question, To what extent did the Cui. subjects transfer learning? If one hypothesizes that they had been taught to subtract with sums to 100 only, one would have to infer the answer to be, To a very great extent. Three fourths of the examples in Set 3 have sums larger than 100, and yet their Ac. mean is 77 percent that of the Tra. sample. Also, 33 of the Cui. subjects made scores of 9 to 12.

Only three classes in the Cui. sample did outstandingly well on Set 3, and these three are among the five classes mentioned in connection with column addition. The number of subjects in these classes is small, only 20. As a group they were very high scorers on the Tra. Test as a whole. Their Ac. mean on this test is 44.50 in comparison with the Ac. mean of 26.22 for the entire Cui. sample. For Set 3 their Ac. mean is 8.75,--3.54 greater than the Ac. mean of the Cui. sample and 1.90 greater than the Ac. mean of the Tra. sample.

The hypothesis that their extraordinary success on Set 3 is to be attributed to an equally extraordinary amount of transfer does not seem to be tenable. A preferred explanation is that they had been taught how to subtract in all the types included in the set. And if these classes had had such "extra instruction,"[1] it is highly probable that in other Cui. classes children were explicitly taught some of the skills in subtracting from three-place sums.

Acceptance of this second explanation does not mean a total denial of transfer, for this notion is implausible. But acceptance does mean a considerable reduction in the estimated degree to which transfer functioned. By how much is not discoverable in the data at hand.

Set 4. Division.
Items 17 - 22, 27 - 31, 45, 48, 49, 61. Table 33

a. The set as a whole. Columns (4) and (7). Set 4 consists of two groups of items. In the first group there are are six items. Two are simple division combinations; four are uneven divisions. Of the latter, three make use of the simple combinations, as in item 20, 16 ÷ 3 = _____, with products less than 34; and one is the example 69 ÷ 10 = _____. Group 2 contains nine division examples, all with digits as divisors, and ranging in difficulty from $2\overline{)48}$ to $5\overline{)987}$.

---

[1] See p. 163 for an explanation of the term "Extra instruction." That Cui. teachers could properly provide "extra" teaching is consistent with the Cui. program as described in Chapter I.

The Ac. means in columns (4) and (7) are 5.95 (Cui.) and 7.80 (Tra.), and the NA means, 4.59 and 3.93, respectively. The difference in Ac. means, in favor of the Tra. subjects, is reliable. Scores of 12 to 15 were made by 39 Tra. subjects and by 22 Cui. subjects; scores of 0 to 3 by 32 Tra. and 47 Cui. subjects.

The Tra. subjects had the greater number of correct answers for 14 of the 15 examples, by 14 or more in 11; and it had the smaller number of incorrect answers in 13, by 9 or more in eight. One Cui. subject and two Tra. subjects stopped all work before attempting answers for item 17, and so, are recorded with NAs for all 15 examples. Six more Cui. subjects and five more Tra. subjects did not reach item 27, and so, are recorded with NAs for the last nine examples. The numbers of NAs because of lack of time mounted rapidly with item 48 (two-thirds of the way through the test) until on item 61 the totals of NAs were 67 (Cui.) and 61 (Tra.).

Nevertheless, the numbers of subjects in the two samples who omitted all items after each of six stopping places are very similar. In only one instance did the numbers differ by as much as 7. In the other five instances the differences

Table 33

The Tra. Test. Scores of Samples in Division

Scores	Cui. Sample						Tra. Sample					
	Items 17 - 22		Items 27 - 31, 45 48, 59, 61		Total Preceding		Items 17 - 22		Items 27 - 31, 45 48, 59, 61		Total Preceding	
	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA	Ac.	NA
(1)	(2)		(3)		(4)		(5)		(6)		(7)	
15					2	2					6	2
14					6	3					13	2
13					9	3					13	2
12					5	2					7	2
11					3	2					6	1
10					5	9					7	1
9			6	18	6	3			8	13	9	6
8			10	3	8	3			12	4	4	7
7			7	5	8	6			14	2	3	5
6	23	3	7	5	7	8	46	8	13	5	7	6
5	17	5	9	2	9	7	13	1	12	2	6	9
4	13	6	6	18	5	16	13	4	8	15	7	12
3	8	9	11	13	10	8	6	6	9	14	5	7
2	15	10	14	20	11	16	12	10	8	27	9	22
1	18	16	12	5	7	8	4	12	11	2	8	4
0	26	71	38	31	19	24	26	79	25	36	10	32
Means	2.89	1.08	3.06	3.48	5.95	4.59	3.66	0.99	4.08	3.01	7.80	3.93
S.D.'s	--	--	3.02	3.13	4.65	4.14	--	--	3.07	2.94	4.99	3.89

The difference in favor of the Tra. group over the Cui. group for accuracy is the only significant difference at the 0.01 level.

amounted to only 1 or 2. On this account, and despite some small degree of resulting inaccuracy, differences in NAs have been, and will be disregarded in comparisons.

The Tra. sample was clearly superior to the Cui. sample in accuracy on the 15 examples in Set 4, and it should have been if the Cui. subjects had learned to divide only two-place products. Seven of the examples in the set have products ranging from 204 to 609.

b. Simple division. Items 17 - 22. Columns (2) and (5). Illustrations of the examples in Group 1 are 20 ÷ 4 and 33 ÷ 6. For this group of six items the Tra. mean for correct answers is 0.77 greater than the Cui. Ac. mean, and the Tra. NA mean is 0.09 smaller than the corresponding Cui. mean. The Tra. subjects made the greater number of correct computations in five examples, by 16 or more in four, and made the smaller numbers of errors in all six examples, by 13 or more in four. Yet, the Cui. sample might have been expected to do as well as did the Tra. sample. If the Cui. subjects had been taught to divide products to 100, they should have learned the simple division combinations and learned how to deal with uneven divisions. And it is precisely these elements of knowledge that were tested in Group 1. The Tra. program, whatever form it took, seems to have been more effective than the Cui. program in teaching knowledge and skill unintentionally common to both programs.

c. Computational division. Items 27 - 31, 45, 48, 59, 61. Columns (3) and (6). There is a very large difference in favor of the Tra. sample between the two Ac. mean for the nine examples in Group 2,--1.02. (The first example in the group is $2\overline{)48}$; the last, $5\overline{)707}$.) The Tra. sample also has the smaller number of NAs by 0.47. Only 11 more Tra. subjects than Cui. subjects made scores of 7 to 9 on Group 2, and 22 fewer Tra. subjects than Cui. subjects made scores of 0 to 2. In every example the Tra. subjects wrote more correct answers, in seven by 14 or more, and smaller numbers of incorrect answers in seven examples, in three by 9 or more.

On the whole, "extra teaching" (the meaning of this term has been fully explained) in some or many Cui. classes is again regarded as the factor chiefly responsible for the ability of a number of Cui. subjects to compute accurately the quotients in the more complicated examples of Set 4. The reader is already familiar with the line of argument. (1) Twenty-two Cui. subjects made scores of 12 to 15. (2) The Ac. mean of the 41 children in the five Cui. classes most successful on Set 4 is 11.1, a mean twice the Ac. mean of the complete Cui. sample and 1.4 times the Ac. mean of the Tra. sample for the division section of their own achievement test. (3) Exclusion of these 41 children from the Cui. sample lowers the Ac. mean of the remaining 79 Cui. subjects to 4.57. It is hardly believable either that the 41 and the 79 children had had the same instruction in division or that, if this be true, the 41 children could transfer so much more of their learning to untaught kinds of division.

This sample of 41 children, which represents a third of the total sample, had capability, however, for more-than-average amounts of transfer, for they were highly proficient in the arithmetic of the Tra. Test. Their Ac. mean on this test is 42.74, in comparison with the Ac. mean of 26.22 for the complete Cui. sample. And it is to be assumed that they profited from their unusual capability through transfer, but improbably enough to explain entirely their great superiority over their fellows in the Cui. sample.

If this argument is sound, then "extra teaching" was chiefly responsible for their demonstrated success in division; and it is not unlikely that "extra teaching" was done in Cui. classes other than the five in question.

Set 5. Multiplication.
Items 44, 47, 57, 62. Table 34.

The Tra. Test contains only four multiplication examples, all in the last third of the instrument. Each example has a three-place numeral as multiplicand (285, 134, 420, and 354) and a digit (5, 6, 3, 4) as multiplier. One example requires renaming in only one column, the hundreds'; the other three require it in all columns.

The Tra. sample has the larger mean for accuracy by 0.62 and the smaller mean for NAs by 0.46. The numbers of correct answers made by the Tra. sample are the larger in all examples, by 15 to 26 in three of them; and the Tra. frequencies of errors are consistently the smaller, but by differences of but 8 or less.

Table 34

The Tra Test.
Scores of Samples in Multiplication

Scores	Items 44, 47, 57, 62			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
4	14	32	23	17
3	7	17	21	16
2	19	26	18	25
1	22	8	20	18
0	58	37	38	44
Means	1.14	1.99	1.76	1.53

Because of the placement of the multiplication items late in the test, there were many omissions, starting with 23 for the Cui. sample and 15 for the Tra. sample in the case of item 44 and reaching 77 and 72 in the case of item 62. At each of the four points where children stopped because of lack of time, the Tra. subjects had the fewer NAs by 5 to 10.

Both because of the small number of examples and because of the extreme frequency of NAs, comparisons of the two samples for proficiency in multiplication are hazardous. All that can be said is that the few data available are consistent with a tentative claim of superiority on the part of the Tra. subjects.

Finally, if the Cui. subjects had been taught to multiply only two-place numerals with products to 100, it is understandable that many of them should have had trouble with the three-place multiplicands in Set 5; but it is less understandable, on this hypothesis, why many of them had no more trouble with the examples than they did. As heretofore, one may challenge the hypothesis of little teaching or offer an explanation in terms of transfer of learning.[2]

Fourteen Cui. subjects made perfect scores of 4, and seven more, scores of 3. Furthermore, the five classes that made the strongest showings in multiplication have Ac. means of 1.7, 1.8, 2.4, 2.5, and 2.6; and the Ac. mean of the 37 children in these five classes is 2.14. This Ac. mean is to be compared with that of the full Cui. sample, 1.14, and with that of the full Tra. sample, 1.76, which had been taught to perform the multiplications tested. The Ac. mean of the 37 children on the Tra. Test is 36.90, or 10.68 greater than that of 26.22 for the whole Cui. sample. This difference amounts to 41.1 percent and demonstrates, on the part of the 37 children, an unusual degree of general arithmetical proficiency as well as their capability of transferring a great deal of their learning. Nevertheless, the "extra teaching" inferred to have been given these classes (and perhaps to other Cui. classes in smaller measure) seems to have been the factor most influential in their success in multiplication.

---

[2] In retrospect the reader may wonder why this problem was not raised in Chapter V where the Tra. program was under study to determine how much ability to transfer learning it produced. The reason the problem was not discussed is that it arose only in a minor way. In connection with three sets in the Cui. Test comments were made to the effect that unsuspected instruction may have been given to some Tra. classes but not to the point of affecting vitally the results of the comparisons. On the other hand, it was known that the Tra. subjects have been taught the skills of computation with abstract numbers beyond the degree of proficiency required by the items in the Cui. Test. The Tra. subjects encountered difficulty at other points in the Cui. Test; e.g., the use of brackets and the meaning of fractions like 3/4 when used in computation.

Set 6. The Meaning of Numbers.
Items 13 - 16, 38 - 42. Table 35.

None of the items in Set 6 calls for computation. All, rather, relate to understandings basic to intelligent computation. Items 12 - 15 and 42 have to do with the structure of numbers; for example, In the number 2,381 the 2 stands for _____, and Write down the number that has 4 tens, 2 units, and 5 hundreds. In item 16, subjects were asked to designate the number "carried" in the example 987 + 71 as 0, 1 unit, 1 ten, or 1 thousand. The remaining four items call for the writing of the smallest or the greatest number under specified conditions; for example, the smallest with the figures 8 4 9 1; the greatest that can be written with three figures.

According to Table 35, the Ac. means of both samples are small (less than a third of the possibility), and that for the Tra. sample is but 0.12 the larger. The Tra. mean for NAs is strikingly the smaller, by 3.56. One Tra. subject and no Cui. subjects failed to reach item 13 within the time limits, and totals of 12 Tra. and 14 Cui. subjects did not reach item 38. In other words, the Tra. sample omitted 69 examples, and the Cui. sample omitted 70 examples because of lack of time. All other NAs were the result of deliberate intent, and the number of such omissions on the part of the Cui. sample is 2.8 times that of the Tra. sample.

The Cui. subjects wrote more correct answers for three items, in one by 18, and the Tra. subjects wrote more correct answers in four, in three by 13 or more. As for errors, the Cui. sample has fewer in two examples, in one by 10; and the Tra. sample, fewer in seven, by 10 or more. On the whole, the two samples made closely comparable records on Set 6; that is, if judged in terms of accuracy, but not in terms of NAs.

The Tra. program, especially when it is taught with intelligent computation as the goal, puts considerable emphasis on the ideas tested in Set 6. Yet, only 25 Tra. subjects made scores of 6 - 9, and this fact, together with their small Ac. mean, seems to imply that in actuality the bulk of them had had little instruction on number meanings as tested here.

Table 35

The Tra. Test.
Scores of Samples on the Meaning of Numbers

Scores	Items 13 - 16, 38 - 42			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
9	2	36	1	5
8	2	27	4	7
7	5	10	10	6
6	8	10	10	2
5	11	6	6	17
4	11	12	9	10
3	18	8	15	10
2	16	3	18	10
1	24	1	23	21
0	23	7	24	32
Means	2.72	6.49	2.84	2.93
S.D.'s	2.32	2.70	2.50	2.79

On the other hand, if one accepts as genuine and fairly comprehensive the description of the Cui. program in Chapter I, Cui. teachers do comparatively little with these ideas, at least through Primary III. That they had done little is borne out by the facts (a) that the Cui. Ac. mean is small (2.72), (b) that only 9 Cui. subjects made scores of 7 - 9, and (c) that there were so many omissions, never less than 59 for an example and as many as 100 for one item: Write down the smallest number that can be written with 5 in tens place. (The totals cited include both deliberate NAs and NAs caused by lack of time.)

The conclusion to which one comes is that there are too many uncertainties in interpreting the data for Set 6 to justify making a confident estimate of the degree to which Cui. subjects transferred understandings to examples not specifically taught.

Set 7. Completing two-operation number sentences. Items 49 - 56. Table 36.

The first four items in this set are of the type 36 + 59 + _____ = 100; the next three of the type, 23 = 7 x 2 + _____; and the last is 28 = 24 - 4 + _____. As the numbers of the items indicate, all in the set appeared in the last fourth of the test. Hence, there were a great many children who did not have the time to make the computations called for. Forty-nine Cui. subjects and 31 Tra. subjects did not even start Set 7, and eight more Cui. and 14 more Tra. subjects stopped before reaching item 53. If to the NAs resulting from lack of time are added intentional omissions, it will be understood why the means for NAs are so great: 4.30 (Cui.) and 4.02 (Tra.) (Table 36).

Table 36

The Tra. Test.
Scores of Samples in Completing Two-Operation Number Sentences

Scores	Items 49 - 56			
	Cui. Sample		Tra. Sample	
	Ac.	NA	Ac.	NA
8	0	42	1	33
7	5	5	6	13
6	7	7	9	4
5	6	6	5	5
4	7	8	5	9
3	7	8	11	8
2	9	4	12	7
1	22	9	18	4
0	57	31	53	37
Means	1.63	4.30	1.87	4.02
S.D.'s	2.17	3.36	2.31	3.34

On the set as a whole the Tra. sample was slightly more accurate than the Cui. sample; but the difference between Ac. means, 0.24, is certainly unreliable. Only four more Tra. than Cui. subjects made scores of 6 to 8. The Tra. subjects have the larger frequencies of correct examples in all eight examples, but by margins as large as 7 and 11 in only two. The Cui. subjects made fewer errors in three examples, by 9, 6, and 1; the Tra. subjects, in four examples, their largest advantages being 8 in one case and 6 in another; but the numbers of errors are large for both samples. For the Cui. sample the numbers of incorrect answers range between 16 and 43; for the Tra. sample, between 22 and 49. As a consequence the frequencies of negative responses (errors plus NAs) are very large, and the frequencies of correct answers are very small. Except for item 49, which was the easiest for both samples, the numbers of correct answers are never larger than 26 for the Cui. sample and 29 for the Tra. sample. Hence, when one says that the two samples were about equal in accuracy, the statement does not mean much.

The items in this set resemble a good many items in the Com. Test in that they are incomplete two-operation number sentences. In the Com. Test the missing numeral is to be inserted in any of the positions in number sentences, initial, intermediate, or final; and no item requires multiplication. In Set 7, three items call for multiplication as the first step in computation; the missing numerals are never in the initial position, and the computations, except in item 49 (9 + 10 + _____ = 25) are more complex than in the Com. Test. Samples are: 36 + 49 + _____ = 100; 243 + 109 + _____ = 461; 29 = 5 x 5 + ____; 28 = 24 - 4 + ____.

Nevertheless, there is enough similarity between the Com. Test items (all acceptable to both the Cui. and the Tra. samples) and the items in Set 7 of the Tra. Test,--

enough similarity to anticipate that subjects successful in the Com. Test should have been successful in Set 7. All the multiplication combinations used in Set 7 (7 x 2, 6 x 3, and 5 x 5) should have been known by the children in both samples; and in the three examples in which they appeared the largest difference between numbers of correct answers is 4. Only one item, 52, made use of three-place numerals, and in this example only two Cui. subjects and nine Tra. subjects secured correct answers. The absence of brackets in the Set 7 items, a factor which was thought to have influenced the Cui. subjects adversely in the Com. Test, may have had this effect also in the Tra. Test. Of course, the Tra. subjects, unlike the Cui. subjects, probably had had some practice with examples exactly like those of Set 7, even if their record for this set does not reveal any real advantage over the Cui. subjects.

To sum up, in view of all the facts mentioned above, one hesitates to assess the extent to which the Cui. subjects profited from transfer of learning. Perhaps a fair estimate would be "somewhat," meaning more than "none" but much less than "considerable" or "substantial."

Concluding Statement

As was true in the case of Chapter VI, so here in this chapter, the attempt to assess the extent of transfer, this time on the part of the Cui. subjects on the Tra. Test, must be reported as having been none too successful. The chief interfering factor was the apparent lack of anything like uniform practice in the Cui. schools. While some diversity was expected, it had been hoped that the Cui. schools as a group had conformed fairly closely to the same arithmetical objectives with respect to learning outcomes in computation anticipated at the end of Grade III.

Miss Law in her description of the Cui. program (Chapter I) warned against putting too much faith in these anticipations, for she made no pretense of outlining the Cui.program, knowing that there is no such program. By contrast the computational objectives of the Tra. program (not the Tra. program) as described by Mr. Allan (also in Chapter I) appear to have set fairly well the pattern of teaching in Tra. schools.

On some sets of skills in the Tra. Test the Cui. subjects did about as well as did the Tra. subjects, despite the fact that the test items were intended to be beyond their capabilities. In some sets, too, the records of the Tra. subjects as a whole were inferior to those of three to five selected Cui. classes.

These findings can be accounted for as the result of transfer of learning on the part of the Cui. subjects. But the amount of transfer required would have had to be very large indeed, so large as to seem incredible. A second explanation is that in some Cui. classes, probably composed of very able children and certainly composed of arithmetically able children, teachers carried instructional skills much further than was expected by either the Cui. or the Tra. panel,--even as far as did the Tra. teachers as a whole. The second explanation is favored by the writer, but with the conviction that transfer also was present and assisted the Cui. subjects in some measure.

The findings of the comparisons of the records of the samples on the seven sets of computational skills in the Tra. Test are summarized below briefly. For the omitted details the reader is referred to appropriate sections in the earlier pages of the chapter.

Set 1. Column addition; 11 items. Illustrations (written horizontally): 26 + 30 + 12 + 31; 493 + 27 + 854; 508 + 5 + 80 + 416. Examples of these kinds, known to have been taught Tra. classes, are absent in the Cui. Test, though simpler kinds appear in the Com. Test, with the additions to be made either vertically or horizontally. The Cui. Ac. mean represents 85.6 percent of the Tra. Ac. mean. While certainly many Cui. subjects had been taught column addition little beyond the addition of two-place addends with sums to 100, there are data interpreted to mean that many others had had instruction that included most if not all the types in Set 1. Transfer of learning is assumed to have occurred rather generally; but its extent cannot be ascertained. In any case its influence is viewed to have been less than was the "extra teaching" mentioned.

Set 2. Horizontal addition; four items. Illustrations are 5 + 4 + 7 + 6 = _____ and 8 + 7 + 0 + 8 + 9 = _____. This set of items should not have been included in the Tra. Test as involving skills possessed only by Tra. subjects. The Com. Test contains one item in horizontal addition (4 + 5 + 6 = _____), a fact which attests the teaching of horizontal addition to Cui. subjects as well as to Tra. subjects. At the worst, Set 2 measured only achievement in both samples. At the best, on the hypothesis that horizontal addition as taught Cui. subjects stopped short of the types in Set 2, these subjects may have had to extend their skill to be successful on the Tra. Test items, and such extension could be called a form of transfer. There is no way to determine the relative validity of the two explanations. Whatever the explanation, the Cui. sample did almost as well as did the Tra. sample on Set 2.

Set 3. Subtraction; 12 items. Illustrations (written horizontally) are: 79 - 35; 135 - 29; 638 - 270; 810 - 79. On the set as a whole the record of the Tra. sample is, as expected, considerably better than that of the

Cui. sample. Even so, the Cui. subjects earned an Ac. mean that is 77 percent that of the Tra. Ac. mean,--and this, on items many of which are about as difficult as can be made with three-place minuends and two- and three-place subtrahends. Again there is evidence that in some Cui. classes subtractions of these kinds, not present in the Com. Test (and so, rejected by the Cui. panel), had actually been taught, with consequent effect on the Cui. Ac. mean. Helpful transfer cannot be ruled out as a factor making for the success of the Cui. sample, but its amount is unknown.

Set 4. Division; 15 items. Illustrative of a group of six items are 20 ÷ 4 = _____ and 69 ÷ 10 = _____; of a group of nine items, $2\overline{)48}$, $3\overline{)343}$, $5\overline{)375}$, and $6\overline{)690}$. The Ac. means of the two samples on the set as a whole are: Cui., 5.95; Tra., 7.80. The difference between these means, in favor of the Tra. subjects, is reliable. Yet, the Cui. subjects' Ac. Mean is 76 percent that of the Tra. sample. "Extra teaching" (in the sense in which that term is here used) seems almost certainly to have been given at least a third of the Cui. subjects; and this fact destroys any chance of estimating the extent to which they profited from transfer of learning.

Set 5. Multiplication; four items. Illustrations (written horizontally) are 5 x 285; 4 x 354. In the Com. Test there is but one computational item in multiplication (Multiply 25 by 4), and this is part of a two-operation example. The inference is that any other multiplication examples proposed by the Tra. panel were rejected by the Cui. panel as inappropriate for Cui. children. This inference is supported by the fact that in the Cui. Test the only multiplications (exclusive of the simple combinations) are 4 x 25 = _____ and 10 x 10 = _____. One is therefore unprepared to discover that for the four difficult multiplication items in the Tra. Test the Cui. mean frequency of correct answers is 65 percent that of the Tra. sample. Three conclusions can be drawn. (a) The Cui. subjects did much better in multiplication than the Cui. panel anticipated. (b) They did so because of unexpectedly large amounts of "extra teaching" with respect to multiplication skills. (c) It is fruitless to conjecture concerning the degree to which the Cui. subjects profited from transfer of learning.

Set 6. Number meanings; nine items. Illustrations are: Put a line under the number that has 8 in tens' place: 873 8,250 386 128; and Write the greatest number that can be written with these figures: 0 9 3 8. Neither sample made a good record on Set 6. The mean frequencies of correct answers for the nine items, 2.72 for the Cui. subjects and 2.84 for the Tra. subjects, represent little more than 30 percent of the possible score. For every item with a correct answer the Tra. subjects omitted another item, and the Cui. subjects omitted 2.4 other items. The fact that each sample outscored the other on three examples seems to indicate that each had the advantage on items that had been specifically taught. As was stated at an earlier point, there are too many uncertainties in interpreting the data for Set 6 to justify making a confident estimate of the degree to which the Cui. subjects transferred understandings to skills not specifically taught.

Set 7. Completing two-operation number sentences; eight items. Illustrations are: 28 + 17 + _____ = 70; 20 = 6 x 3 + _____; 28 = 24 - 4 + _____. The record of the Tra. sample in terms of mean number of correct answers is 15 percent better than that of the Cui. sample; but this advantage means little because of the excessive frequencies of omissions, both those that were deliberate and those caused by lack of time. For every example correctly done the Cui. subjects omitted 2.9 other examples and the Tra. subjects, 2.1 other examples. As can be seen in the examples cited above, the items in Set 7 are written without ( )'s or [ ]'s. Similarly written examples in the Com. Test seemed to confuse some of the Cui. subjects. In the absence of these, to them, familiar cues, they did not seem to know where to start the computations. For them, in Set 7 negative transfer may well have offset part of their positive transfer, the extent of which, on balance, has been previously described by the term "somewhat."

In a word, in the results for not a single one of the seven sets of items in the Tra. Test is it possible to find convincing evidence bearing on the extent of transfer of learning in the performance of the Cui. subjects. That transfer of learning actually occurred is fully conceded. Indeed, its occurrence is positively asserted. It is only the amount of transfer which is in question, and this could not be ascertained, or even guessed.

PART IV

CONCLUSIONS AND IMPLICATIONS

CHAPTER VIII

SUMMARY AND FINDINGS, WITH RELATED RESEARCH

Purpose of the Study

One purpose of the investigation--the practical purpose--was to compare the effectiveness of two programs for teaching Scottish children to acquire competence in arithmetical computation. The programs in question are the Cuisenaire (Cui.) system of instruction and what is here called the Traditional (Tra.) system,--both systems as they were taught in the years 1963-1965, and not as now taught. These two programs have been fully described in Chapter I. The Cui. program is taught to a very limited extent in American schools, and the Tra. program resembles somewhat that taught in this country thirty years ago.

The second purpose of the study, also practical in its own way, was to try out the usefulness of what is believed to be a unique design for evaluative research like that in this inquiry. The rationale of the design and the procedure it entails will be described below in the section entitled The Research Design.

Two bases were employed to compare the effectiveness of the two programs in developing computational proficiency. One was achievement on skills known to have been taught in both Cui. and Tra. Scottish schools; the other basis was performance on skills not taught, in order to assess the relative extent to which the programs engendered concepts and understandings that made for fruitful transfer of learning.

Procedure

Subjects

One group of subjects consisted of 539 Scottish children who had studied arithmetic according to the Cui. program for three years, in the grades Primary I, Primary II, and Primary III. A second group of 570 Scottish children had been taught arithmetic in the same grades according to the Tra. program. Extreme care was exercised to make sure that the subjects in each group had had contact only with the one program to which they were assigned.

The two large groups of subjects were closely comparable with respect to chronological age, to amounts of instructional time per week over the three-year period, and to scores on an intelligence test. They were drawn from 18 Cui. classes and from 18 Tra. classes in 12 Cui. and 13 Tra. schools, distributed over a 50-mile band in Scotland from Aberdeen at the northeast to Ayrshire at the southwest. School officials and teachers, as well as pupils, were uniformly cooperative and to an extent that is rare in research of this kind.

Tests and testing

A Scottish intelligence test, the Moray House 8+ Reasoning Test, was recommended by Dr. A. E. G. Pilliner, then Director of the Godfrey Thomson Research Unit for Educational Research, of the University of Edinburgh, and was administered to all subjects. The pupils' tests were scored and normed especially for the total number of research subjects by the staff of this Unit.

Three computational tests, prepared by panels of Scottish experts in the two programs, were tried out several times and were altered until they served the intended purposes and could be given in 30 minutes of working time. The three tests are designated as the Com. (Common) Test, the Cui. Test, and the Tra. Test, and each has a reliability coefficient greater than 0.90 obtained by the test-retest method. The research testing was done by eleven girls in the graduating class of Moray House College of Education, who also scored the tests.

The intelligence test and the Com. Test were administered to all subjects in the morning and in the afternoon of the same day. The Cui. classes took the Cui. Test and then the Tra. Test in the morning and in the afternoon of the second day, and the Tra. classes took the Tra. Test and then the Cui. Test in the morning and in the afternoon of the second day. The reason for the change in order of the last two arithmetic tests will be made clear shortly.

The Research design

Quite unintentionally, one may be sure, in evaluative studies of the kind here attempted, the findings of the test used are commonly predetermined by the nature of the test itself. That is to say, the test gives the advantage to one of the competing groups. This happens when the test is suited to the objectives of one of the programs under study. The consequence is that, while the favored group is measured with respect to achievement, the disadvantaged subjects are tested on untaught subject matter and are not tested on the subject matter peculiar to their program. A serious effort was made to avoid this source of error.

A panel of three Scottish experts in the Cui. program first prepared a test which was intended to comprehend all computational skills taught in Cui. schools in the first three grades. Another panel of Scottish experts in the Tra. program independently constructed an equally comprehensive test for that program. Next, the two tests were compared item by item, and those items which appeared in both tests were used, perhaps in modified form, to make up the Com. Test. Under this plan, the Com. Test measured achievement on the part of both groups.

The Cui. panel then prepared a test consisting of skills peculiar to the Cui. program, resulting in the Cui. Test. In similar fashion the Tra. panel made up a test of skills supposedly taught only in the Tra. program. For the Cui. subjects the Cui. Test was an achievement test, but for the Tra. subjects it afforded opportunity to reveal how well they could transfer learned skills and understandings to untaught skills. In the same manner the Tra. Test was an achievement test for the Tra. subjects, but a transfer test for the Cui. subjects. In the scheduling of the testing, each group of subjects took the transfer test last so as to prevent undue frustration on all preceding tests.

This research design has two significant advantages. (a) It assured fairness to both Cui. and Tra. subjects in the measurement of achievement and should yield an answer to the question, Which program, if either, was superior to the other in developing competence in computation involving skills taught in the programs? (b) It enabled children in each program to reveal accomplishments of use in connection with skills outside the limits of that program and should supply an answer to the question, To what extent did each program inculcate ideas, understandings, and the like making for transfer to untaught skills?

Measures obtained

For each of the arithmetic tests as wholes scores of accuracy and of rate of work were obtained for both groups. These gross scores for the 539 Cui. subjects and for the 570 Tra. subjects were studied in various ways through the analysis of variance, (Chapter III), and the results will be included in the Summary of Findings below.

Data on accuracy (frequencies of correct answers), errors, and non-attempts (NAs) were found for each item in each of the three tests, but for reduced samples of 120 Cui. and 120 Tra. subjects constituted by the use of a table of random numbers. These samples proved to be comparable in all important respects to the original total groups from which they had been drawn. The results of the item analyses will be incorporated in the Summary of Findings below.

Reliability and educational significance

When appropriate, the reliability of differences between means was subjected to the t-test of significance, and differences at the 0.01 level were accepted as reliable. This procedure was employed in the analysis of variance in Chapter III and also in the item analyses (Chapters V - VII) when sets or groups of items numbered 10 or more.

Yet, a difference may be "significant at the 0.01 level" (a statistical concept) and not educationally significant, for the difference between means may be so small as to be of no importance in the practical enterprise of educating children. Hence, means had to differ by 10 percent of the total number of items in the test (an arbitrary standard), in order to be accepted as educationally significant.

In the item analyses of 120 randomly selected tests from each group, the differences between numbers of correct answers or errors or non-attempts had to satisfy two criteria in order to be regarded as reliable and educationally significant. To be said to be superior on a set or group of items one sample had to outscore the other both consistently and by margins of reasonable size. The standard employed for the second criterion is arbitrary once again,--10 percent. That is to say, on a given example 12 more members of one sample than of the other had to have correct answers.

For a fuller statement of the rationale of the procedures stated briefly above, the reader is referred to pages 75-78 of Chapter IV.

Summary of Findings: Relative Achievement

Analysis of variance (Chapter III)

No evidence is presented in Chapter III to support a claim that either of the arithmetic programs under investigation was superior to the other. At the outset the groups of subjects in the two programs were determined to be comparable in intelligence and chronological age; and the amount of time given to arithmetic instruction in the two groups was found to be the same.

An examination of the results on the test which contains only items approved both by the Cui. and by the Tra. panel (the Com. Test) revealed no significant differences between the mean scores of the two total groups. In both programs subjects attempted about the same number of items; they correctly computed answers for approximately the same number of examples; and the ratios of the numbers of items completed satisfactorily to the numbers of items attempted were nearly the same. All differences on the Com. Test favor the Cui. subjects; but no differences are statistically significant, and most certainly are not of a magnitude to be considered as having educational significance.

Scores of sub-samples classified according to treatment group, intelligence, and the length of arithmetic instructional time suggest one clear advantage for the Cuisenaire program. Children in the Cui. group classified as Low in intelligence and Long in instructional time scored significantly higher on the Com. Test than any other sub-sample classified as Low in intelligence. The differences between this sub-sample and all other Low intelligence sub-samples surpassed the 10 percent minimum in raw-score accuracy which had been established as a basis for judging a result to be of educational significance. This result is all the more impressive when one considers that the Long instructional period of the Cui. sub-sample averaged four minutes per day less than the Long period of the Tra. sub-sample.

When arithmetic achievement (instead of intelligence), time, and treatment group were the criteria for classifying subjects, no differences between sub-samples were great enough to be educationally significant when instructional time was analyzed as the source of variation. However, the subjects assigned to the Long instructional period in the Low intelligence Cui. sub-sample scored significantly higher than their counterparts in the Short instructional period. This result gives further credence to the superiority of the Cui. program for some children.

From the evidence summarized relevant to the achievement of the children in the two instructional programs, it is reasonable to suggest that children identified as low in intelligence (and perhaps achievement) and exposed to a relatively long period of instruction in arithmetic will gain more through involvement in the Cuisenaire program. No claims for the superiority of either program can be made with respect to other sub-samples or the total samples selected for this investigation.

Item analysis (Chapter V)

The major conclusion reached in the study of test scores by analysis of variance--to the effect that the total groups of Cui. and Tra. subjects were equally competent in computing answers for examples both groups had been taught--is strongly supported by the findings of item analysis.

In this latter form of attack 120 subjects were drawn from each of the original large Cui. and Tra. groups by means of a table of random numbers. As shown in Table 10 these samples were closely comparable in all respects examined, to the groups from which they were taken. The test papers for the samples, and not for the original groups, were scrutinized item by item.

The 72 items of the Com. Test were classified into seven sets of comparatively homogeneous items on the basis of the skill or skills needed. These seven sets, sometimes divided into sub-groups in recognition of distinguishable differences in their mathematical requirements, are listed below together with illustrative examples.

Set 1. The simple number combinations
9 + 8 = _____
15 - 7 = _____

Set 2. Supplying numerals in number sentences
2 x _____ = 12
_____ ÷ 2 = 7
18 = 7 + _____ + 6
_____ = 13 + 31
64 - _____ = 11

Set 3. Horizontal and/or vertical addition
19 + 26 + 34 = _____ **or**

$$\begin{array}{r} 19 \\ 26 \\ +\ 34 \\ \hline \end{array}$$

Set 4. Fractions Multiply $\frac{1}{2}$ of 16 by 3. _____
What is the fifth part of 30? _____

Set 5. Structure and divisibility of numbers

Write this number in figures. Four hundred and ten._____
Write three numbers that divide into 24 exactly.

_____ _____ _____

Set 6. Two-operation examples 34 - 25 + 8 = _____

Multiply 10 by 5 and take away 4. _____

Set 7. Time measures; Scottish money

1 hour - $\frac{1}{2}$ hour = _____ minutes
1s. 6d. + 2s. 9d. = _____s. _____d.

In each set as a whole, judged primarily by frequencies of correct and of incorrect answers, the two samples were a match for each other.

(a) In every set, except Set 3, each sample had the larger number of correct answers in one or more examples. (Table 18)

(b) In every set, except Set 3, each sample had the fewer errors in one or more examples. (Table 19)

(c) Of the whole 72 items in the Com. Test, only eight are "critical examples," six in favor of the Cui. sample, two in favor of the Tra. sample. (Table 21)

The peculiar status of Set 3 in (a) and (b) and the meaning of finding (c) call for comment. But, first, it should be noted that findings (a) and (b) point to inconsistency rather than consistency in both strength and weakness in the work of the samples within sets. And consistency is one of the two criteria employed to determine superior performance significant for educational practice. The other criterion is that differences be of reasonable size. This criterion was not satisfied either to any extent, as the reader can decide for himself by citing the tables mentioned above or the detailed statements of results in Chapter V.

Now, as for the peculiar status of Set 3: the items in this set gave all subjects the choice between adding numerals horizontally or in columns, and oral directions before the testing and printed statements on the test blanks indicated this fact. Yet, on a substantial number of Cui. papers there is unmistakeable evidence of confusion: the subjects in question simply did not understand what to do in the presence of the two forms in which each example was printed. That they were capable of adding horizontally is attested by the fact that the Cui. panel approved of the Set 3 items for the Com. Test and by the further fact that Cui. subjects did well on such items in the Tra. Test. That they were able to add numerals in columns is attested by their later success with examples of this kind in the Tra. Test. Hence, the apparent weakness of the Cui. subjects on the Com. Test on items 33, 24, and 62 - 64 is to be discounted as having been produced by an extraneous factor; namely, inability to decide what to do, and not inability to perform the computations.

As for finding (c): a "critical example" is here defined as one in which one sample secured 12 or more correct answers than did the other sample; or, stated differently, an example in which 12 more subjects in one sample than in the other computed accurately. This difference of 12 represents 10 percent of the total sample, and thus satisfies the arbitrary criterion of "reasonable size" mentioned above.

To conclude: the evidence as a whole is interpreted to be very convincing that the Cui. subjects and the Tra. subjects, whether considered as total groups by an analysis of variance or as samples by item analysis. were a match for each other on the Com. Test. This is to say that they were equally competent in the computational skills tested as those skills were represented in the examples chosen for that test. And, accordingly, one can say that when computational competence means competence on these skills, neither the Cui. program nor the Tra. program is the more effective in developing proficiency in computation.

As was reported in Chapter VII, some of the Cui. classes had had more than the anticipated amount of instruction on computational skills. The effect was to make the Com. Test easier for the Cui. sample as a whole than it otherwise would have been and to increase the size of the Ac. means for this sample. The fact that the Tra. sample still compared favorably with the Cui. sample on the Com. Test should remove any doubt concerning the effectiveness of the Tra. program.

The conclusion of equal effectiveness of the programs may not hold for the totality of computational proficiency; for example, in the case of skills taught in one program but not in the other. Nor does the conclusion necessarily hold for groups of children taught exclusively according to the Cui. program or according to the Tra. program, say, for six years, and tested in Primary VI instead of in Primary III. (However, as a matter of fact, this last possibility is purely academic, at least in Scotland, for exceedingly few Scottish Cui. schools continue the Cui. program beyond Primary III.)

Summary of Findings: Transfer of Learning

It would be pleasant indeed to be able to report that the Cui. sample transferred learning to the extent of 39 percent and the Tra. sample, to the extent of 21 percent, or *vice versa*. No such neat and exact findings were anticipated for reasons that will appear in the discussion of the Tra. performance on the Cui. Test and of the Cui. performance on the Tra. Test. The simple truth is that means are not available to arrive at such precise mathematical expressions of amounts of transfer. Even more fundamental reasons will be advanced in the section of the report entitled Theoretical Considerations.

The two bodies of information concerning transfer gleaned from the sub-sample analyses (a) of scores made by Tra. subjects on the Cui. Test and (b) of scores made by Cui.

subjects on the Tra. Test will be combined and presented below under the Analysis of Variance. The findings of the item analyses in (a) and in (b) which were presented in Chapters VI and VII will be reported separately.

Analysis of variance

Normally, one would expect subjects in the Cui. instructional group to perform at a higher level than subjects in the Tra. group on the test designed for the Cui. program. Likewise, the corresponding expectation would hold for the Tra. subjects on the Tra. Test. The level of performance of children in each instructional program on the test designed for the other program may indicate a degree of transferability. One notable exception merits comment. Those subjects participating in a Long instructional time period who were identified as Low intelligence in the Cui. treatment group scored considerably higher on the Tra. Test than Low intelligence subjects exposed to a Short instructional time. This is the only instance in which a subsample of children in one treatment group scored higher than its counterpart in the other treatment group on a test designed for the other group.

For some reason, which is not hypothesized at this time, the children classified as Low in intelligence who studied arithmetic each day for a relatively Long period of time in the Cui. program appear to gain more both in achievement and in transfer than do children classified by any other set of criteria employed in this study. There is good reason to believe that the Cui. program is especially effective for children with a relatively low intelligence level if adequate time is provided for the teaching-learning process.

The Tra. subjects on the Cui. Test

Item analysis. (Chapter VI) The Cui. Test contains computational items involving skills thought by the Cui. panel to have been taught to all Cui. subjects but to no Tra. subjects. The degree of success of the Tra. subjects on these items was intended to indicate the extent to which they were able to transfer skills to examples unfamiliar to them.

The Ac. mean of the Tra. sample on the Cui. Test is 16.85, and the Ac. mean of the Cui. sample on the same test is 21.90. The Tra. subjects, then, computed accurately 74.6 percent as well as did the Cui. subjects. If the Tra. sample on this test had to work only examples totally new to them, they must have transferred learning to a huge extent; either this, the explanation lies in the influence of other unexpected factors.

Below are listed the sets of items into which the 60-item Cui. Test was divided, mainly on the basis of their mathematical requirements.

Set 1. One-operation examples 27 - 6 = _____ 5 x 6 = _____
32 ÷ 8 = _____ 3 x 20 = _____

Set 2. Fraction computations and two-operation examples with fractions and whole numbers
3/5 of 10 = _____ 1/24 of 48 = _____
10 ÷ 2 = 4 + _____ 1/8 of 16 = $\frac{1}{2}$ of _____
Find the half of 96.

Set 3. Computation with one and two brackets, fractions and whole numbers
[6 x 3] ÷ 9 = _____
[3/4 of 16] ÷ [14 - 12]= _____

Set 4. Number meanings; doubling numbers
Double 250 _____
Put X on the 4 that means 4 tens. 4 4 4

Set 5. Number progressions; comparative size of numbers
20, 18, 16, 14, _____
Write down these numbers in order of greatness.
10 60 30 100 80

Set 6. Two- and three-operation examples with one or two sets of brackets
12 -[$\frac{1}{4}$ of _____] = 9
[5 + 6] x [14 - 12] = _____

Set 7. Computations with Scottish money
£ 3/4 = £$\frac{}{20}$ £1$\frac{1}{2}$ = _____ shillings

Set 8. Measurement items
1/8 of a gallon + 2 pints = _____ pints
How many inches in 3 feet? _____

Some instruction not known by the Cui. panel to have been given at least part of the Tra. subjects and the relatively poor showing of the Cui. sample,--these two factors make it hazardous to estimate the degree to which the Tra. subjects transferred learning in Set 1.

In dealing with Set 2, a good many Tra. subjects correctly computed answers for examples with unit fractions by converting the examples into "short divisions," and transferred the skill in the latter to good effect in items like 1/3 of 9 = 9 ÷ _____.

Set 3 items containing only one bracket and small numbers with simple computations (e.g., 4 + [$\frac{1}{2}$ of 8] = _____, and [6 x 3] ÷ 9 = _____) were worked successfully by many Tra. subjects in spite of their lack of experience with brackets. However, few of these subjects were successful on more complicated items.

Set 4 provided no clear evidence on the problem of extent of transfer, for the Cui. subjects surpassed the Tra. subjects, and vice versa, according to the type of example they had been taught. The few items on number meanings might well have been excluded from the Cui. Test.

The items in Set 5 relating to the comparative size of numbers had been taught to both Tra. and Cui. subjects and should probably not have been included in the test. In dealing with number proportions, not emphasized in their program, the Tra. subjects transferred ideas concerning number relationships acquired from such activities as counting by 5's.

The poor showing of the Cui. subjects in Set 6 made the record of the Tra. subjects, also poor, look much better than it actually was. Tra. subjects gave some evidence of transferring understanding of computational procedures.

In set 7 the Tra. panel performed fairly well. Though on the whole they had had little instruction in computing with Scottish money, they had been taught the values of the coins and some of them were able to transfer their knowledge of computational procedures.

The results of Set 8, like those of Set 4, are ambiguous respecting the extent of Tra. transfer. The Tra. subjects had been taught the units of linear, volume, and avoirdupois; and knowledge of these units and their relationships was of course basic to computations as in the Set 8 items; but how much more the Tra. subjects had been taught is uncertain.

<ins>To conclude</ins>: Two obstacles interfered in the attempt to discover the extent to which the Tra. sample profited from transfer of learning in the computations called for in the Cui. Test. One was the fact that in some instances the Tra. subjects may have been taught skills that (the Cui. panel assumed) they were not to have been taught. The result was a combination of directed learning and of transfer, a combination that defied analysis to differentiate the effects of either from the other.

The other obstacle was that the Cui. Test in several of its parts was too difficult even for the Cui. subjects. In another place the suggested explanation was that the Cui. subjects tested were, in general, inferior in computational competence to those used in standardizing the test. But, be this as it may, the consequence was that the Cui. record, being none too good at points, set a relatively low standard for comparison with the Tra. record and therefore made the showing of the Tra. subjects seem better than it may have been.

The Tra. subjects unquestionably transferred learning even in the most troublesome sets. (See Theoretical Considerations below for the defense of this statement.) The trouble was, however, that its occurrence and its extent were not as readily identifiable as could have been wished. Some evidence of transfer (but not how much) has been noted in the paragraphs above for Sets 2, 3, 5 (number progressions only), 6, and possibly 7.

The Cui. Subjects on the Tra. Test

Item analysis. (Chapter VII) The records of the Cui. and Tra. samples on the Tra. Test, specifically designed for children who had studied according to the Tra. program, were compared in order to determine the extent of transfer on the part of the Cui. subjects in computing answers for examples not supposedly taught to them.

The records of the Cui. and Tra. samples on the Tra. Test, specifically designed for children who had studied arithmetic according to the Tra. program, were compared in order to determine the extent of transfer on the part of the Cui. subjects in computing answers for examples assumedly not taught them.

On the Tra. Test the Ac. means of the two samples are: Cui., 26.22; Tra., 32.26. Therefore, the Cui. subjects did 81.3 percent as well as did the Tra. subjects. The performance of the Cui. sample is far superior to any that could have been anticipated if they had not been taught the needed skills, as the Tra. panel supposed. On the assumption that they had not received such instruction, they must have made extraordinary use of transfer; or else there must have been one or more unsuspected factors operating in their behalf. The only other explanation which comes to mind amounts to a denial of the validity of the assumption just mentioned.

The Tra. Test has 63 items. These items were divided into seven sets. Within each set the items were as homogeneous as practicable,--that is, homogeneous with respect to the skills needed. These sets are named, and illustrative items are now provided.

Set 1. Column addition.

$$\begin{array}{r} 26 \\ 30 \\ 12 \\ +\ 31 \\ \hline \end{array} \qquad \begin{array}{r} 526 \\ 863 \\ +345 \\ \hline \end{array} \qquad \begin{array}{r} 493 \\ 27 \\ 854 \\ +\ \ 75 \\ \hline \end{array}$$

Set 2. Horizontal addition

6 + 4 + 7 + 6 = _____

8 + 7 + 0 + 8 + 9 = _____

Set 3. Subtraction

$$\begin{array}{r} 79 \\ -\ 35 \\ \hline \end{array} \qquad \begin{array}{r} 400 \\ -\ 35 \\ \hline \end{array} \qquad \begin{array}{r} 905 \\ -\ 369 \\ \hline \end{array} \qquad \begin{array}{r} 511 \\ -\ \ 31 \\ \hline \end{array}$$

Set 4. Division

20 ÷ 4 = _____ $2\overline{)48}$ $3\overline{)343}$ $5\overline{)987}$
32 ÷ 5 = _____

Set 5. Multiplication $\begin{array}{r}285\\ \underline{\times\quad 5}\end{array}$ $\begin{array}{r}134\\ \underline{\times\quad 6}\end{array}$ $\begin{array}{r}354\\ \underline{\times\quad 4}\end{array}$

Set 6. Number meanings

Put a line under the number that has exactly 40 tens.

40 400 04 2,040

Write down the smallest number that can be written with these figures. 8 4 9 1

Set 7. Completing two-operation number sentences

342 + 109 + _____ = 461 20 = 5 x 5 - _____

Considerable space will be taken to interpret the results on Set 1, because the situation uncovered in this case recurs in the case of other sets as well. On Set 1 the Ac. means are: Cui., 6.62; Tra., 7.73. The Cui. Ac. mean is 85.6 percent that of the Tra. sample, a remarkable showing on the part of the Cui. sample if one hypothesizes that these subjects had not been taught to compute answers for examples like those in Set 1. But there is ample reason to doubt the validity of this hypothesis.

Thirty-nine Cui. children made scores of 9 to 11 on Set 1, and 11 is the highest possible score. Moreover, the 35 children in the five Cui. classes with the best records on Set 1 have an Ac. mean of 9.06, which is 2.44 points greater than the Ac. mean of the whole Cui. sample and 1.33 points greater than the Ac. mean of the Tra. sample.

The 35 children cited were far above average in computational skill on the Tra. Test as a whole. Their Ac. mean of 37.10 is to be compared with the Cui. sample's Ac. mean of 26.22 and with the Tra. sample's Ac. mean of 32.26. One could argue that, being so capable, they could have transferred learning to the extent necessary to achieve their high standing in column addition. This explanation seems unacceptable to the writer as calling for an incredibly large amount of transfer,--incredibly large even to one who sets great store by transfer. A more credible explanation in his view is that these particular Cui. classes (and perhaps other Cui. classes in smaller measure) had been taught the skills required in Set 1.[1]

This last explanation does not deny the strong probability of some transfer, but it greatly reduces the reliance to be placed upon transfer alone in accounting for the excellence of the Cui. sample in column addition. Further, it justifies the conclusion that the data for this skill are such as to negate any attempt to assess the extent of the role of transfer in Set 1. Now for the findings in other sets.

---

[1] That competence in arithmetical computation need not make for large amounts of transfer is illustrated by the records of one Cui. class of 29 members, the class which has the highest Ac. mean for the Cui. Test of all the Cui. classes tested,--49.90, to be compared with the Cui. sample's Ac. mean of 21.90. On the Com. Test their Ac. mean is 60.24, almost 20 points greater than the Ac. mean of the entire Cui. sample. Yet, on the Tra. Test their Ac. mean is but 27.66, only 1.44 points above the Cui. sample's Ac. mean. Expressed in percentages, their Ac. means of 83.2 on the Cui. Test and of 83.7 on the Com. Test dropped to 43.9 on the Tra. Test. This illustration, in the writer's opinion, weakens the case for transfer of learning as the sole, or even the major explanation for the high Cui. accuracy on Set 1.

In Set 2 there is no way to differentiate between transfer and directed learning. Both samples had been instructed in horizontal addition. As evidence, one such item appears in the Com. Test; hence, it had to be approved by both the Cui. and the Tra. panels.

It is impossible to estimate the extent of Cui. transfer in Set 3. The situation with regard to subtraction is the same as that described above for Set 1.

The same statement applies to Set 4, division.

It also applies to Set 5, multiplication.

In Set 6 poor records were made by both samples, each excelling on the items that had been taught it.

In Set 7, the Cui. subjects were handicapped by the absence of brackets in incomplete number sentences, a condition which was also troublesome to them in the Com. Test. Their poorer record in Set 7 was not therefore caused by inability to make the necessary computations.

To conclude: In not one of the seven sets was it possible to secure evidence concerning the extent of transfer in the Cui. sample's performance on the Tra. Test.[2]

---

[2] For detailed data, going far beyond the brief statements above regarding each set, the reader is referred to Chapter VII.

(That there was transfer is assumed.) In four sets (1, 3, 4, and 5) the reason is the unexpectedly large amount of instruction given Cui. subjects on skills assumed not to have been taught. However unfortunate this circumstance may be for the purpose of this inquiry, no complaint can be lodged against the teachers who gave this instruction. They clearly were entitled to do what they did, and what they did is not inconsistent with the Cui. program, which is flexible. In the description of the program in Chapter I no restricions are placed on the limits to which teaching may carry children. On the contrary, variation in classroom practice is there said to be common in Cui. schools,--commoner in them, probably, than in Tra. schools where the program tends to be more uniform.

Theoretical considerations

The terms transfer, memory, and learning are customarily used without clear differentiation, as if everyone understands what each term means and how it is related to or is separate from the others. Transfer and learning refer to processes and contain the idea of action,--one learns and one transfers learning. On the other hand, memory refers to a static condition, and behavior related to it is denoted by such verbs as recall and recognize.

Actually, the distinction between transfer on the one hand and recall and recognition on the other is arrived at arbitrarily. As a matter of fact, one rarely, if ever, recalls or recognizes a learned item in a situation which is identical with the situation in which it was learned. Rather, there are always (or almost always) differences between the situations,--differences within the organism, if for no other reason than the effects of prior uses of the given item, and differences outside the organism in the context in which the learned item recurs. And these differences imply that the learner does more than merely to recall or recognize: he transfers his learning.[3]

---

[3] The writer is grateful to his colleague, Dr. Arthur Jensen, for the formulation of his solution of the problem.

> Here is the arbitrary, operational distinction I would make between behavior classified as memory and behavior classified as transfer.
>
> First, all transfer involves memory (that is, the persistence of the effects of prior learning), but memory does not necessarily imply transfer.
>
> The essential difference can be pointed out in terms of the following paradigm:
>
> | Original Learning | Memory or Transfer |
> |---|---|
> | $S - R$ | $S^1 - R$ |
>
> If S^1 is discriminably different from S to the subject, I would call the response (R) an instance of transfer. If S^1 is not discriminable from S to the subject, I would call the response an instance of memory.
>
> By discriminable I mean only that the stimuli differ enough for the subject to be able to learn a discrimination such as
>
> $S - R_1$
>
> $S^1 - R_2$.

It is for this reason that the phrase "no transfer" does not appear in this manuscript. The possibility of "no transfer" may be real, but it is, in the writer's opinion, exceedingly uncommon, especially in the learning of structured subject matter like mathematics. The problem then is not whether transfer occurs--it must occur if there is to be learning--but rather the extent of transfer in any instance of learning.

Review of Related Research

A diligent search failed to uncover much published research in which the comparative merits of the Cui. and the Tra. programs are assessed.[4] Those found, together with an unpublished inquiry will be reviewed. In addition, other articles of interest will be referred to. Out of courtesy, a beginning will be made with Scottish reports.

In a series of three articles in successive issues of The Scottish Education Journal in 1958 Karatzinas and Renshaw report on their "Primary Arithmetic Inquiry." In the first two articles (references 2 and 3 in the bibliography at the end of the chapter) they summarize "Teachers'

---

[4] In the references here reviewed, with one exception, the term traditional is not used to designate the program and the subjects contrasted with the Cui. program and subjects. The commonest word employed is control.

Views of the Cuisenaire Method," based on information gained by a questionnaire addressed to "forty teachers who started to use the Cuisenaire materials at the beginning of the 1957-58 session in Edinburgh Corporation schools, class Primary I."

The first of the 22 questions illustrates the types of question asked, and the method reporting the quantitative data is typical.

> "Do you consider that by using the Cuisenaire material you have achieved better results (so far) than you might have achieved in the same time without the material?

Reply	Yes	No	Not sure	No answer	Total
Number	27[1]	4	8[2]	1[3]	40

The raised numerals after each Number refer to interpretations and explanations and well considered judgments condensed into a paragraph. These comments add greatly to the value of the summaries and make the articles well worth reading by persons interested in the Cui. materials and their use. Yet, one must remember that the respondents had taught with the materials for but a single year, and it is generally recognized that it takes years to make a teacher really expert in directing learning with the Cui. apparatus.

The third article (4), one page in length, reports the results of an experiment undertaken "to ascertain by means of tests the effectiveness over the first eighteen months of primary schooling of arithmetic teaching conducted with the aid of the [Cui.] material, compared with teaching without the material."

The subjects in the Cui. group consisted of a class of forty boys in one school; the subjects in the Non-Cui. group, of a class of 14 girls and 24 boys in another school. The schools involved were selected by a local school official in such a manner as to make groups of children comparable with respect to potential and to "the educational stimulus provided both at home and school." Again, the strong probability is that the program in the Non-Cui. school resembled that in the Tra. schools in the present study.

In intelligence the two groups were closely similar as shown by scores on two intelligence tests, one a picture test. To measure arithmetical competence six parts of the Schonell Diagnostic Arithmetic Test were used. In the four parts entitled Addition, Subtraction, A to E - Addition, and A to D - Subtraction, the means of the two groups were so nearly alike that no difference was reliable at the 5 percent level. Two other parts of the test--A to K - Multiplication, and A to K - Division--could not be given the Non-Cui. subjects who "had not made sufficient progress to be able to attempt" these tests. The investigators point out as highly significant the fact that the Cui. group did very well with division and multiplication, at the same time that it maintained parity with the Non-Cui. group in the other two operations.

Dr. T. Renshaw has also supplied the writer with a few unpublished data on the relative competence of 26 Cui. and 26 Non-Cui. subjects as demonstrated on a comprehensive test of 103 items. All subjects were starting their first term in the secondary school and centered around 13 years of age. It is not clear whether the Cui. subjects had studied mathematics for six years exclusively according to an extended Cui. program; and the other group, Non-Cui., may or may not have pursued what is named the Tra. program in the writer's investigation, but it probably did so because of its general popularity.

According to his letter, Dr. Renshaw's research "was an enquiry made incidentally during the try-out" of a new mathematics test. On this account, the comparability of the two groups was less than he hoped for. The Non-Cui. group averaged about six months the older, but had a slightly smaller I.Q. on tests given years before. On the other hand, at the end of Primary VI the two groups earned very nearly the same score on a standardized mathematics test.

It was believed that the test data on the try-out test would reveal "strong differences that might have persisted beyond the primary stage of schooling." On this test the mean scores of the two groups were: Cui., 41.5; Non-Cui., 42.9, with respective S.D.'s of 10.7 and 12.9. Fifteen items were identified in which one group surpassed the other by reliable or near-reliable differences in frequencies of correct answers. Nine of the comparisons were in favor of the Cui. group.[5]

Lucow has three articles reporting an experimental study made under the auspices of the Manitoba Teachers' Society in schools in cities and towns outside of Winnepeg, Canada. The mean age of the children, who were just entering Grade 3, was just over eight years, both in the Cui. group and in the control group, the latter children being taught according to a prescribed system resembling the Tra. program. Reference 5 was not available to the writer. Reference 6 presents data which are reported in reference 7. Hence, the information supplied in this last reference furnishes the basis of the review of the Lucow investigation.

The subject matter taught was restricted to multiplication and division. The Cui. children had had considerable instruction on these topics in grades 1 and 2; the Non-Cui. children had had none. Consequently, it was decided to make the critical measures those, not of absolute gains in achievement, but of relative growth during the experimental period. By general consent the cooperating teachers concentrated the presentation of the two topics in six weeks instead of the three or four months usually taken. By means of an ingenious method of grouping schools into "blocks" and by using the statistical technique of analysis of covariance, Lucow sought to deal with several troublesome problems. One block of subjects comprised 66 subjects (30 boys, 36 girls); the other block, 63 subjects (31 boys, 32 girls).

---

[5] If, as seems likely to the writer in view of the situations he found in Cui. schools,--*if* the Cui. subjects had actually followed the Cui. program for, say, three years only and were then changed to the Tra. program, the results in the comparisons above are about what would be expected.

Lucow summarizes his findings in terms of levels of significance of differences between the mean scores of Cui. and control subjects, first, for each block of three schools as a whole, then for each separate school. In one block, 11 differences are significantly in favor of the Cui. children at the 0.01 level, and in four more at the 0.05 level; in nine instances there were no significant differences. In the second block the Cui. subjects have five differences significant at the 0.01 level and two more at the 0.05 level, leaving 17 unreliable differences.

The investigator concluded that ". . . the Cuisenaire method proved effective in teaching third grade arithmetic." However, ". . . other current methods of instruction also proved effective." "There is some evidence that the Cuisenaire method operates better in a rural setting . . . than in an urban setting, also with high-IQ and middle-IQ children in a rural setting, but not much better with low-IQ children." "Urban children [except in one school] learned just as well under any method at all levels of intelligence."

"There is only a slight indication that girls take to the Cuisenaire method better than boys."

Passy (8) has reported his findings for some 1800 third grade children on the arithmetic sub-tests of the Stanford Achievement Test, Elementary Battery. Three samples of subjects were used: 990 that had been taught with the aid of the Cui. materials; 375 subjects taught with non-Cui. materials; and 500 that had had the benefit of "pre-Cuisenaire materials." The three samples were comparable in intelligence, according to scores on the California Test of Mental Maturity, Short Form, and according to socio-economic status, according to scores on the Hamburger Rating Scale.

The two systems of instruction compared were the Cui. program and another described as a "meaningful program." Passy's data indicate that children utilizing the Cui. materials achieved "significantly less at the 5% level of significance" on the arithmetic tests, though the average score of each sample was at or above grade level.

Unquestionably the most impressive, comprehensive, and thorough-going investigation of the relative effectiveness of different programs of arithmetic instruction has been published by Biggs (1). At the time he made his study he was associated with the National Foundation for Educational Research in England and Wales, located in London, England. At present he is Research Officer for Mathematics in Monash University, Victoria, Australia.

Two of the programs in his inquiry were the Cui. and the Tra. programs. Unfortunately his findings cannot be compared with those in this investigation, for his subjects were enrolled in the third year of English junior schools (roughly, grade 5). But one of his conclusions is well worth quoting:

> Using conventional problem and mechanical tests[6] as criteria, there is no evidence that the use of uni-model structural materials, such as the Cuisenaire or the Stern materials, will produce results with average children that differ from those obtained under traditional methods, in similar school conditions.

In view of the excellence of the Biggs investigation--and his report--one hesitates to raise a critical question. Perhaps, though, the writer may be excused if he asks just one question. In the writer's study it was exceedingly difficult to find schools in which the Cui. program had been followed in Scottish schools for three years. Biggs's subjects had been in school two years longer. During those years the cooperating schools in the writer's study would almost without exception have been taught the Tra. program. There is a strong possibility that the same thing happened in the Biggs Cui. schools. If so, his subjects labelled as Cui. subjects represented the effects of a combination of both the Cui. and the Tra. programs, with no way of separating out the effects peculiar to either program.

---

[6] Mechanical tests: tests of computational skills with abstract numbers.

There is the probability that one or more published research reports pertinent to the present investigation may have been overlooked. Certainly in the search for relevant research no attempt was made to collect unpublished reports, of which there may be many. The writer knows that in one school system, that of Vancouver, B.C., the practice has been followed of releasing on occasion mimeographed accounts of informal studies of the Cui. program which has been taught there for quite some time. Not being at hand, the Vancouver papers have not been included in this review of research; but they should be available to interested readers.

Critique of the Design of the Investigation

In the writer's opinion, despite the practical limitations cited earlier and to be cited again shortly, the design of the investigation is theoretically sound. In the application of the design in this instance it did not prove to be as useful as it might well have been under other conditions.

It seems to be essential in an evaluative study such as that here reported, to measure achievement by means of a test which is fair to all subjects regardless of the particular instructional program they have pursued. The Com. Test was constructed in a manner intended to serve this purpose. However, the purpose was not fully realized, for in some Cui. classes computational skills had been taught considerably beyond the limits anticipated by the Cui. panel. Lacking this information, the members of this panel rejected as too difficult items proposed by the Tra. panel, with the consequence that the Com. Test may well have favored the Cui. subjects,--by how much is unknown.

Likewise, it seems essential in an evaluative study like this one, to broaden the concept of computational competence to include ideas and understandings which can function in the case of unfamiliar computational skills,--this through the transfer of learning. The Cui. Test, prepared by the Cui. panel and containing skills peculiar to the Cui. program, was expected to measure the extent of transfer manifested by the Tra. subjects in taking this test. Correspondingly, the Tra. Test was expected to measure transfer on the part of the Cui. subjects. These tests did not yield the results hoped for.

The Cui. Test proved to be unduly difficult for the Cui. sample of subjects. On an instrument which measured achievement for them, they earned a mean Ac. score representing but 36.5 percent of the possibility; and on some sets of items they did even more poorly. The effect was unfortunate so far as the purposes of the investigation are concerned. The results of the Cui. subjects on the Cui. Test had been planned to provide the basis for estimating the extent of transfer on the part of the Tra. subjects; but the low "standard" set by the Cui. sample made the poorer showing of the Tra. sample look better than it should have appeared, and would have appeared on a test better adapted to the abilities of the Cui. subjects.[7]

The Tra. Test failed to measure the extent of transfer on the part of Cui. subjects for quite a different reason, the reason mentioned above; namely, the unanticipated amount of instruction on computational skills in many Cui. classes. Hence, it was impossible to determine how much of the success of the Cui. sample was the consequence of direct teaching, and how much, the result of transfer.

The design adopted for this inquiry can be productive of definitive findings concerning the relative effectiveness of differing instructional programs only when those programs are considerably more "standardized" than they were in this study. Yet, the cost to be paid for this "standardization" may be too costly in terms of the most desirable forms of teaching and learning; and the writer hesitates to propose anything of the kind, however helpful the "standardization" might be for evaluative research.

---

[7] The reader may see in the comments concerning the Cui. and the Tra. Tests criticism of the panels which constructed them. No criticism is intended, and none is warranted. Lack of uniformity in objectives of arithmetic teaching among the schools committed to each program, but especially among those committed to the Cui. program, presented serious obstacles to each panel in selecting test items that would be "fair" to the particular group of children concerned. The best the panels could do was to make judgments of what they hoped to be "typical" of classroom practice in Cui. or Tra. schools.

On the other hand, "standardization" has already proceeded rather far in American schools, at least in the teaching of arithmetic. Education in this country, regrettably, is still too largely textbook education. On this account, the results of arithmetic instruction in a group of schools using textbook series A might be compared with the results in another group of schools using textbook series B. In this case, the research design tried out in the Scottish schools in this investigation might prove to be of value.

Chapter References

1. Biggs, J. B. *Mathematics and the Conditions of Learning.* Slough, Bucks, England: National Foundation for Educational Research in England and Wales, 1967. pp. xi + 441.

2. Karatzinas, D., and Renshaw, D. "Teachers' Views of the Cuisenaire Method," Part I. *The Scottish Educational Journal*, September 19, 1958. P. 575.

3. ______. "Teachers' Views of the Cuisenaire Method," Part II. *Ibid.*, September 26, 1958. Pp. 595-596.

4. ______. "Testing the Effectiveness of the Cuisenaire Material," Part III. *Ibid.*, October 3, 1958. P. 613.

5. Lucow, William H. "Cuisenaire Method Compared with Current Methods of Teaching Multiplication and Division." *The Manitoba Teacher*, September-October, 1962. (Report not available.)

6. ______. "Testing the Cuisenaire Method," *The Arithmetic Teacher*. November, 1963. Pp. 435-438.

7. ______. "An Experiment with the Cuisenaire Method in Grade 3." *American Educational Research Journal*, May, 1964. Pp. 159-167.

8. Passy, Robert A. "The Effect of Cuisenaire Materials on Reasoning and Computation." *The Arithmetic Teacher*, November, 1963. Pp. 439-440.

APPENDIX

....

COM. TEST

Sc. ....

Ra. ....

Your name ....

Your teacher's name ....

Your school ....

Write the missing numbers on lines like this:

2 + 8 = ... 10 − 4 = ...

7 + ... = 8 2 × .... = 6

.... + 2 = 3 .... − 4 = 2

Add either way, across or up and down; but just one way.

2 + 2 + 1 = ... *or*

$$\begin{array}{r} 2 \\ 2 \\ +\ 1 \\ \hline \\ \hline \end{array}$$

Write out your working if you wish.

1. 8 + 5 = ...

2. 16 ÷ 2 = ...

3. 17 − 9 = ...

4. 2 × 9 = ...

5. 6 + 7 = ...

6. 12 ÷ 3 = ...

7. 5 × 7 = ...

8. 14 − 9 = ...

9. 15 ÷ 3 = ...

10. 3 × 8 = ...

11. 9 + 8 = ...

12. 15 − 7 = ...

---

13. 2 × .... = 12

14. 12 − .... = 7

15. 6 + ... = 14

16. 9 − .... = 0

17. .... + 7 = 10

18. .... − 5 = 7

19. .... ÷ 2 = 7

20. .... + 8 = 8

---

21. 12 + ... = 20

22. 4 + 5 + 6 = ...

23. 9 + 3 + ... = 16

24. 18 = 7 + ... + 6

25. .... + 16 = 24

26. Add 20 to ½ of 12. ....

27. Find ½ of 20 and divide by 5. ....

28. Multiply ½ of 16 by 3. ....

---

29. 37 = 25 + ...

30. 16 + ... = 29

31. .... = 14 + 21

32. 3 dozen − 30 = ...

Add either way.

33. 19 + 26 + 34 = ... *or*

$$\begin{array}{r} 19 \\ 26 \\ +\ 34 \\ \hline \\ \hline \end{array}$$

34. 25 + 37 + 8 = ... *or*

$$\begin{array}{r} 25 \\ 37 \\ +\ 8 \\ \hline \\ \hline \end{array}$$

---

35. What is the half of 16 ? ....

36. What is the half of 86 ? ....

37. What is a quarter of 24 ? ....

---

38. 5d. + 5d. = ...s. ....d.

39. 9d. − 5d. = ...s. ....d.

40. 8d. × 2 = ...s. ....d.

41. ½ of 10d. = ...s. ....d.

42. 6d. + 6d. + 2d. = ...s. ....d.

43. 1s. 1d. − 8d. = ...s. ....d.

44. ¼ of 8d. = ...s. ....d.

45. £14 + £6 + £8 + £2 = £..........

Write these numbers in figures

46. Three hundred and forty-five. ....

47. Six hundred and five. ....

48. Four hundred and ten ....

---

49. 12 + 26 − 30 = ...

50. 34 − 25 + 8 = ...

51. 66 − 30 − 8 = ...

52. 23 + 41 − 32 = ...

---

53. Find 2 times 8 and add 10. ....

54. Multiply 10 by 5 and take away 24. ....

55. What is a fifth part of 30 ? ....

---

56. 35 − 18 = ...

57. 46 − 17 = ...

58. 64 − .... = 11

59. 33 − .... = 28

60. .... − 13 = 13

61. .... + 15 = 51

Add either way.

62. 65 + 44 + 8 = ... *or*

$$\begin{array}{r} 65 \\ 44 \\ +\ \ 8 \\ \hline \\ \hline \end{array}$$

63. 49 + 56 + 84 = ... *or*

$$\begin{array}{r} 49 \\ 56 \\ +\ 84 \\ \hline \\ \hline \end{array}$$

64. 75 + 8 + 49 = ... *or*

$$\begin{array}{r} 75 \\ 8 \\ +\ 49 \\ \hline \\ \hline \end{array}$$

---

65. 2s. 2d. − 1s. 5d. = ...s. ....d.

66. $\frac{1}{5}$ of 1s. 8d. = ...s. ....d.

67. 1s. 6d. + 2s. 9d. = ...s. ....d.

68. 1 hour − $\frac{1}{2}$ hour = ... minutes

69. 50 minutes and 20 minutes will make

.... hours .... minutes.

---

70. Which of these numbers divide into 12 exactly ? Put X on them.

7 4 6 5 3

71. Write three numbers that divide into 20 exactly.

....

72. Write three numbers that divide into 24 exactly.

....

....

CUI. TEST

Sc. ....

Ra. ....

Your name ....

Your teacher's name ....

Your school ....

Write the missing numbers on lines like this:

2 + 8 = ... 10 − .... = 5

2 + ... = 5 .... − 4 = 1

Write out your working if you wish.

(Tra. classes only.)

In sums like these, first do the work in the [ ]'s.

1. 4 + [2 × 1] = ...
   2 × 1 = ...

2. [10 ÷ 2] − 1 = ...
   10 ÷ 2 = ...

3. .... = 12 − [2 × 3]
   2 × 3 ....

4. [$\frac{1}{2}$ of 10] = [2 + ...]
   $\frac{1}{2}$ of 10 = ...

5. [9 − 1] + [6 − 2] = ...
   9 − 1 = ... 6 − 2 = ...

6. [2 × 6] = [3 × 2] + ...
   2 × 6 = ... 3 × 2 = ...

1. $27 - 6 =$ ....

2. $5 \times 6 =$ ....

3. $8 \times 8 =$ ....

4. $5 \times 10 =$ ....

5. $3 \times 20 =$ ....

6. $32 \div 8 =$ ....

7. $10 \times 10 =$ ....

8. $4 \times 25 =$ ....

9. $90 \div 3 =$ ....

10. $\frac{1}{3}$ of $12 =$ ....

11. $\frac{2}{3}$ of $9 =$ ....

12. $\frac{3}{5}$ of $10 =$ ....

13. $\frac{1}{4}$ of $16 =$ ....

14. $\frac{1}{24}$ of $48 =$ ....

15. Find the half of 96. ....

16. What fraction of 6 is 2 ? ....

17. What fraction of 12 is 4 ? ....

18. What fraction of 20 is 15 ? ....

19. $\frac{1}{3}$ of $9 = 9 \div$ ....

20. $8 \times 4 = 16 \times$ ....

21. $\frac{8}{8}$ of $8 = 0 +$ ....

22. $\frac{1}{8}$ of $16 = \frac{1}{2}$ of ....

23. $10 \div 2 = 4 +$ ....

24. $4 + [\frac{1}{2}$ of $8] =$ ....

25. $[6 \times 3] \div 9 =$ ....

26. $[40 + 4] - [10 + 4] =$ ....

27. $[2 \times 4] - [4 \times 2] =$ ....

28. $[\frac{3}{4}$ of $8] + [2 + 3] =$ ....

29. $[\frac{4}{5}$ of $10] - [\frac{1}{2}$ of $6] =$ ....

30. $[\frac{3}{4}$ of $16] \div [12 - 6] =$ ....

31. Double 48. ....

32. $100 - 50 =$ ....

33. What number comes next after 99 ? ....

34. Put X on the 4 that means 4 tens. 4 4 4

35. Double 250. ....

Write down the number that comes next.

36. 5, 10, 20, 40, ...

37. 20, 18, 16, 14, ...

38. Fill in the missing numbers.

4, 8, ..., 32, ..., 128

39. Write down these numbers in order of greatness.

1, 5, 3, 6, 4

....

40. 10, 60, 30, 100, 80

....

41. $[5 \times 2] = [3 \times 3] +$ ....

42. $12 - [\frac{1}{2}$ of ....$] = 9$

43. $[6 \times 6] = [5 \times 2] +$ ....

44. $[\frac{1}{3}$ of $18] = 10 -$ ....

45. $[\frac{3}{4}$ of $100] - 5 =$ ....

46. $[4 \times 5] \div [5 \times 4] =$ ....

47. $[5 + 6] \times [14 - 12] =$ ....

---

48. $\frac{1}{2}$ of .... = 4d.

49. $\frac{5}{12}$s. + $\frac{1}{4}$s. = ... pennies.

50. £$1\frac{1}{2}$ = ... shillings.

51. £$\frac{3}{4}$ = £$\frac{\dots\dots\dots\dots}{20}$

52. What fraction of a pound is 2s. ? ....

53. 3s. + 3s. + $\frac{1}{2}$ of £1 = ...s.

---

54. How many inches in 3 feet ? ....

55. 1 foot − $\frac{1}{2}$ foot = ... inches.

56. How many pints in 2 gallons ? ....

57. $\frac{1}{8}$ of a gallon + 2 pints = ... pints.

58. $1\frac{1}{2}$ pounds = ... ounces.

59. What must you add to 24 to get 63 ? ....

60. How much greater is $\frac{1}{2}$ of 24 than $\frac{1}{4}$ of 16 ?

…………………………………………………………

TRA. TEST

Sc. ……………………………

Ra. ……………………………

Your name ………………………………………………………………

Your teacher's name ……………………………………………………

Your school ……………………………………………………………

Write the missing numbers on lines like this:

$1 + 8 =$ ………… $\qquad 10 -$ ………… $= 5$

$3 +$ ………… $= 5 \qquad$ ………… $- 2 = 2$

Write out your working if you wish.

In this test, division sums are written like this: $4\overline{)48}$

If this is the way you write division sums, put the answer above the line, like this: $\begin{array}{r} 12 \\ 4\overline{)48} \end{array}$

If you write division sums like this, $4\,\underline{)48}$ you may change the division sums and write your work like this: $\begin{array}{r} 4\,\underline{)48} \\ 12 \end{array}$

1.	2.	3.	4.	5.	6.
Add	**Add**	**Add**	**Add**	**Add**	**Add**
26	30	46	20	42	98
30	4	73	43	50	67
12	21	9	32	87	76
+ 31	+ 13	+ 28	+ 94	+ 69	+ 18

7.	8.	9.	10.	11.	12.
Subtract	**Subtract**	**Subtract**	**Subtract**	**Subtract**	**Subtract**
79	94	86	135	115	400
− 35	− 69	− 49	− 29	− 68	− 35

---

13. **Put a line under the number that has 8 in the tens' place.**

873 8,250 386 128

14. **In the number 2,381 the 2 stands for** ....

15. **Put a line under the number that has exactly 40 tens.**

40 400 04 2,040

16. 987 + 71 **Put a line under the number you must carry in working this sum.**

0 1 unit 1 ten 1 thousand

Divide

17. 24 ÷ 3 = ...

18. 20 ÷ 4 = ...

19. 32 ÷ 5 = ...

Divide

20. 16 ÷ 3 = ...

21. 33 ÷ 6 = ...

22. 69 ÷ 10 = ...

Add

23. 5 + 4 + 7 + 6 = ...

24. 9 + 0 + 8 + 3 + 7 = ...

Add

25. 4 + 2 + 9 + 7 + 5 = ...

26. 8 + 7 + 0 + 8 + 9 = ...

---

27. **Divide**

$$2\overline{)\,48}$$

28. **Divide**

$$4\overline{)\,84}$$

29. **Divide**

$$4\overline{)\,204}$$

30. **Divide**

$$3\overline{)\,343}$$

31. **Divide**

$$2\overline{)\,416}$$

---

32. **Add**

$$\begin{array}{r} 576\\ 863\\ +\ \ 345\\ \hline \\ \hline \end{array}$$

33. **Add**

$$\begin{array}{r} 493\\ 27\\ +\ \ 854\\ \hline \\ \hline \end{array}$$

34. **Add**

$$\begin{array}{r} 493\\ 735\\ +\ \ 180\\ \hline \\ \hline \end{array}$$

35. **Subtract**

$$\begin{array}{r} 860\\ -\ \ 428\\ \hline \\ \hline \end{array}$$

36. **Subtract**

$$\begin{array}{r} 905\\ -\ \ 369\\ \hline \\ \hline \end{array}$$

37. **Subtract**

$$\begin{array}{r} 511\\ -\ \ 39\\ \hline \\ \hline \end{array}$$

---

38. Write down the **smallest** number that can be written with **5** in the **tens'** place. ....

39. Write down the **greatest** number that can be written with these figures: 0 9 3 8. ....

40. Write down the **smallest** number that can be written with these figures: 8 4 9 1. ....

41. What is the **greatest** number that can be written with three figures? ....

42. Write down the number that has **4 tens, 2 units,** and **5 hundreds.** ....

---

43. **Subtract**

$$\begin{array}{r} 638\\ -\ \ 270\\ \hline \\ \hline \end{array}$$

44. **Multiply**

$$\begin{array}{r} 285\\ \times\ \ 5\\ \hline \\ \hline \end{array}$$

45. **Divide**

$$3\overline{)\,375}$$

46. **Add**

$$\begin{array}{r} 493\\ 27\\ 854\\ +\ \ 75\\ \hline \\ \hline \end{array}$$

47. **Multiply**

$$\begin{array}{r} 134\\ \times\ \ 6\\ \hline \\ \hline \end{array}$$

48. **Divide**

$$5\overline{)\,987}$$

GO ON TO THE NEXT PAGE

Write the missing numbers.

49. 9 + 10 + ... = 25

50. 28 + 17 + ... = 70

51. 36 + 49 + ... = 100

52. 243 + 109 + ... = 461

53. 23 = 7 × 2 + ...

54. 20 = 6 × 3 + ...

55. 29 = 5 × 5 + ...

56. 28 = 24 − 4 + ...

57.	58.	59.	60.	61.	62.	63.
Multiply	**Subtract**	**Divide**	**Add**	**Divide**	**Multiply**	**Subtract**
420	810	6) 690	508	5) 707	354	700
× 3	− 79		5		× 4	− 680
			80			
			+ 416			

D.P., Edin.—1,550—7/66